

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 April 2001 (19.04.2001)

PCT

(10) International Publication Number  
**WO 01/27857 A2**

(51) International Patent Classification<sup>7</sup>: **G06F 19/00**

(21) International Application Number: **PCT/US00/28413**

(22) International Filing Date: 13 October 2000 (13.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/159,176	13 October 1999 (13.10.1999)	US
60/217,658	10 July 2000 (10.07.2000)	US
60/217,251	10 July 2000 (10.07.2000)	US
09/663,968	19 September 2000 (19.09.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/159,176 (CIP)
Filed on	13 October 1999 (13.10.1999)
US	60/217,658 (CIP)
Filed on	10 July 2000 (10.07.2000)
US	09/663,968 (CIP)
Filed on	19 September 2000 (19.09.2000)
US	60/217,251 (CIP)
Filed on	10 July 2000 (10.07.2000)

(71) Applicant (for all designated States except US): **SEQUENOM, INC.** [US/US]; 11555 Sorrento Valley Road, San Diego, CA 92121-1331 (US).

(72) Inventors: and

(75) Inventors/Applicants (for US only): **BRAUN, Andreas** [DE/US]; 11237-6 Carmel Creek Road, San Diego, CA 92130 (US). **KÖSTER, Hubert** [DE/US]; 8636-C Via Malkona Drive, La Jolla, CA 92037 (US). **VAN**

**DEN BOOM, Dirk** [DE/DE]; Eppendorfer Weg 205 D, D-20253 Hamburg (DE). **PING, Yip** [US/US]; 3641 Copley Avenue, San Diego, CA 92116 (US). **RODI, Charlie** [US/US]; 13823 Recuerdo Drive, Del Mar, CA 92014 (US). **HE, Liyan** [CN/US]; 10948 Creek Bridge Place, San Diego, CA 92128 (US). **CHIU, Norman** [CA/US]; 1128 Caminito Alvarez, San Diego, CA 92126 (US). **JURINKE, Christian** [DE/DE]; Rombergstrasse 22, 20255 Hamburg (DE).

(74) Agents: **SEIDMAN, Stephanie, L.** et al.; Heller Ehrman White & McAuliffe, Suite 700, 4250 Executive Square, La Jolla, CA 92037 (US).

(81) Designated States (national): AE, AG, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING POLYMORPHIC GENETIC MARKERS**

(57) Abstract: Process and methods for creating a database of genomic samples from healthy human donors, methods that use the database to identify and correlate polymorphic genetic markers and other markers with diseases and conditions are provided.

WO 01/27857 A2

## METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING POLYMORPHIC GENETIC MARKERS

### RELATED APPLICATIONS

Benefit of priority to the following applications is claimed herein:

- 5 U.S. provisional application Serial No. 60/217,658 to Andreas Braun, Hubert Koster; Dirk Van den Boom, filed July 10, entitled "METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING POLYMORPHIC GENETIC MARKERS"; U.S. provisional application Serial No. 60/159,176 to Andreas Braun, Hubert Koster, Dirk Van den Boom, filed October 13, 1999, entitled "METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING POLYMORPHIC GENETIC MARKERS"; U.S. provisional application Serial No. 60/217,251, filed July 10, 2000, to Andreas Braun, entitled "POLYMORPHIC KINASE ANCHOR PROTEIN GENE SEQUENCES, POLYMORPHIC KINASE ANCHOR PROTEINS AND METHODS OF DETECTING POLYMORPHIC KINASE ANCHOR PROTEINS AND NUCLEIC ACIDS ENCODING THE SAME"; and U.S. application Serial No. 09/663,968, to Ping Yip, filed September 19, 2000, entitled "METHOD AND DEVICE FOR IDENTIFYING A BIOLOGICAL SAMPLE."

- Where permitted that above-noted applications and provisional applications are incorporated by reference in their entirety.

### FIELD OF THE INVENTION

- Process and methods for creating a database of genomic samples from healthy human donors. Methods that use the database to identify and correlate with polymorphic genetic markers and other markers with diseases and conditions are provided.

### BACKGROUND

- Diseases in all organisms have a genetic component, whether inherited or resulting from the body's response to environmental stresses, such as viruses and toxins. The ultimate goal of ongoing genomic research is to use this information to develop new ways to identify, treat and potentially cure these diseases. The first step has been to screen disease tissue and identify genomic changes at the level of individual samples. The identification of these "disease"

-2-

markers has then fueled the development and commercialization of diagnostic tests that detect these errant genes or polymorphisms. With the increasing numbers of genetic markers, including single nucleotide polymorphisms (SNPs), microsatellites, tandem repeats, newly mapped introns and exons, the challenge to the medical and pharmaceutical communities is to identify genotypes which not only identify the disease but also follow the progression of the disease and are predictive of an organism's response to treatment.

Currently the pharmaceutical and biotechnology industries find a disease and then attempt to determine the genomic basis for the disease. This approach is time consuming and expensive and in many cases involves the investigator guessing as to what pathways might be involved in the disease.

#### Genomics

Presently the two main strategies employed in analyzing the available genomic information are the technology driven reverse genetics brute force strategy and the knowledge-based pathway oriented forward genetics strategy. The brute force approach yields large databases of sequence information but little information about the medical or other uses of the sequence information. Hence this strategy yields intangible products of questionable value. The knowledge-based strategy yields small databases that contain a lot of information about medical uses of particular DNA sequences and other products in the pathway and yield tangible products with a high value.

#### Polymorphisms

Polymorphisms have been known since 1901 with the identification of blood types. In the 1950's they were identified on the level of proteins using large population genetic studies. In the 1980's and 1990's many of the known protein polymorphisms were correlated with genetic loci on genomic DNA. For example, the gene dose of the apolipoprotein E type 4 allele was correlated with the risk of Alzheimer's disease in late onset families (see, *e.g.*, Corder *et al.* (1993) *Science* 261: 921-923; mutation in blood coagulation factor V was associated with resistance to activated protein C (see, *e.g.*, Bertina *et al.* (1994) *Nature* 369:64-67); resistance to HIV-1 infection has been shown in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene (see,

-3-

*e.g.*, Samson *et al.* (1996) *Nature* 382:722-725); and a hypermutable tract in antigen presenting cells (APC, such as macrophages), has been identified in familial colorectal cancer in individuals of Ashkenzi jewish background (see, *e.g.*, Laken *et al.* (1997) *Nature Genet.* 17:79-83). There may be more than three million polymorphic sites in the human genome. Many have been identified, but not yet characterized or mapped or associated with a marker.

#### Single nucleotide polymorphisms (SNPs)

Much of the focus of genomics has been in the identification of SNPs, which are important for a variety of reasons. They allow indirect testing (association of haplotypes) and direct testing (functional variants). They are the most abundant and stable genetic markers. Common diseases are best explained by common genetic alterations, and the natural variation in the human population aids in understanding disease, therapy and environmental interactions.

Currently, the only available method to identify SNPs in DNA is by sequencing, which is expensive, difficult and laborious. Furthermore, once a SNP is discovered it must be validated to determine if it is a real polymorphism and not a sequencing error. Also, discovered SNPs must then be evaluated to determine if they are associated with a particular phenotype. Thus, there is a need to develop new paradigms for identifying the genomic basis for disease and markers thereof. Therefore, it is an object herein to provide methods for identifying the genomic basis of disease and markers thereof.

#### SUMMARY

Databases and methods using the databases are provided herein. The databases comprise sets of parameters associated with subjects in populations selected only on the basis of being healthy (*i.e.*, where the subjects are mammals, such as humans, they are selected based upon apparent health and no detectable infections). The databases can be sorted based upon one or more of the selected parameters.

The databases are preferably relational databases, in which an index that represents each subject serves to relate parameters, which are the data, such as age, ethnicity, sex, medical history, etc. and ultimately genotypic information,

-4-

that was inputted into and stored in the database. The database can then be sorted according to these parameters. Initially, the parameter information is obtained from a questionnaire answered by each subject from whom a body tissue or body fluid sample is obtained. As additional information about each sample is obtained, this information can be entered into the database and can serve as a sorting parameter.

The databases obtained from healthy individuals have numerous uses, such as correlating known polymorphisms with a phenotype or disease. The databases can be used to identify alleles that are deleterious, that are beneficial, and that are correlated with diseases.

For purposes herein, genotypic information can be obtained by any method known to those of skill in the art, but is preferably obtained using mass spectrometry.

Also provided herein, is a new use for existing databases of subjects and genotypic and other parameters, such as age, ethnicity, race, and gender. Any database can be sorted according to the methods herein and alleles that exhibit statistically significant correlations with any of the sorting parameters can be identified. It is noted, however, is noted, that the databases provided herein and randomly selected databases will perform better in these methods, since disease-based databases suffer numerous limitations, including their relatively small size, the homogeneity of the selected disease population, and the masking effect of the polymorphism associated with the markers for which the database was selected. Hence, the healthy database provided herein, provides advantages not heretofore recognized or exploited. However, the methods provided herein can be used with a selected database, including disease-based databases, with or without sorting for the discovery and correlation of polymorphisms. In addition, the databases provided herein represent a greater genetic diversity than the unselected databases typically utilized for the discovery of polymorphisms and thus allow for the enhanced discovery and correlation of polymorphisms.

The databases provided herein can be used for taking an identified polymorphism, and ascertaining whether it changes in frequency when the data is sorted according to a selected parameter.

-5-

One use of these methods is correlating a selected marker with a particular parameter by following the occurrence of known genetic markers and then, having made this correlation, determining or identifying correlations with diseases. Examples of this use are p53 and Lipoprotein Lipase polymorphism.

- 5 As exemplified herein, known markers are shown to have particular correlation with certain groups, such as a particular ethnicity or race or one sex. Such correlations will then permit development of better diagnostic tests and treatment regimens.

- 10 These methods are valuable for identifying one or more genetic markers whose frequency changes within the population as a function of age, ethnic group, sex or some other criteria. This can allow the identification of previously unknown polymorphisms and ultimately a gene or pathway involved in the onset and progression of disease.

- 15 The databases and methods provided herein permit, among other things, identification of components, particularly key components, of a disease process by understanding its genetic underpinnings and also permit an understanding of processes, such as individual drug responses. The databases and methods provided herein also can be used in methods involving elucidation of pathological pathways, in developing new diagnostic assays, identifying new potential drug  
20 targets, and in identifying new drug candidates.

The methods and databases can be used with experimental procedures, including, but are not limited to, *in silico* SNP identification, *in vitro* SNP identification/verification, genetic profiling of large populations, and in biostatistical analyses and interpretations.

- 25 Also provided herein, are combinations that contain a database provided herein and a biological sample from a subject in the database, and preferably biological samples from all subjects or a plurality of subjects in the database. Collections of the tissue and body fluid samples are also provided.

- 30 Also, provided herein, are methods for determining a genetic marker that correlates with age, comprising identifying a polymorphism and determining the frequency of the polymorphism with increasing age in a healthy population.

-6-

Further provided herein are methods for determining whether a genetic marker correlates with susceptibility to morbidity, early mortality, or morbidity and early mortality, comprising identifying a polymorphism and determining the frequency of the polymorphism with increasing age in a healthy population.

5 Any of the methods herein described can be used out in a multiplex format.

Also provided are an apparatus and process for accurately identifying genetic information. It is another object of the herein that genetic information be extracted from genetic data in a highly automated manner. Therefore, to  
10 overcome the deficiencies in the known conventional systems, a method and apparatus for identifying a biological sample is proposed.

Briefly, the method and system for identifying a biological sample generates a data set indicative of the composition of the biological sample. In a particular example, the data set is DNA spectrometry data received from a mass  
15 spectrometer. The data set is denoised, and a baseline is deleted. Since possible compositions of the biological sample may be known, expected peak areas may be determined. Using the expected peak areas, a residual baseline is generated to further correct the data set. Probable peaks are then identifiable in the corrected data set, which are used to identify the composition of the  
20 biological sample. In a disclosed example, statistical methods are employed to determine the probability that a probable peak is an actual peak, not an actual peak, or that the data too inconclusive to call.

Advantageously, the method and system for identifying a biological sample accurately makes composition calls in a highly automated manner. In  
25 such a manner, complete SNP profile information, for example, may be collected efficiently. More importantly, the collected data is analyzed with highly accurate results. For example, when a particular composition is called, the result may be relied upon with great confidence. Such confidence is provided by the robust computational process employed .

30 DESCRIPTION OF THE DRAWINGS

-7-

Figure 1 depicts an exemplary sample bank. Panel 1 shows the samples as a function of sex and ethnicity. Panel 2 shows the caucasians as a function of age. Panel 3 shows the Hispanics as a function of age.

Figures 2A and 2C show an age- and sex-distribution of the 291S allele of the lipoprotein lipase gene in which a total of 436 males and 589 females were investigated. Figure 2B shows an age distribution for the 436 males.

Figure 3 is an exemplary questionnaire for population-based sample banking.

Figure 4 depicts processing and tracking of blood sample components. Figure 5 depicts the allelic frequency of "sick" alleles and "healthy" alleles as a function of age. It is noted that the relative frequency of healthy alleles increases in a population with increasing age.

Figure 6 depicts the age-dependent distribution of ApoE genotypes (see, Schächter *et al.* (1994) *Nature Genetics* 6:29-32).

Figure 7A-D depicts age-related and genotype frequency of the p53 (tumor suppressor) codon 72 among the caucasian population in the database. \*R72 and \*P72 represent the frequency of the allele in the database population. R72, R72P, and P72 represent the genotypes of the individuals in the population. The frequency of the homozygous P72 allele drops from 6.7% to 3.7% with age.

Figure 8 depicts the allele and genotype frequencies of the p21 S31R allele as a function of age.

Figure 9 depicts the frequency of the FVII Allele 353Q in pooled versus individual samples.

Figure 10 depicts the frequency of the CETP (cholesterol ester transfer protein) allele in pooled versus individual samples

Figure 11 depicts the frequency of the plasminogen activator inhibitor-1 (PAI-1) 5G in pooled versus individual samples

Figure 12 shows mass spectra of the samples and the ethnic diversity of the PAI-1 alleles.

Figure 13 shows mass spectra of the samples and the ethnic diversity of the CETP 405 alleles.

Figure 14 shows mass spectra of the samples and the ethnic diversity of the Factor VII 353 alleles.

Figure 15 shows ethnic diversity of PAI-1, CETP and Factor VII using the pooled DNA samples.

5        Figure 16 shows the p53-Rb pathway and the relationships among the various factors in the pathway.

Figure 17, which is a block diagram of a computer constructed to provide and process the databases described herein, depicts a typical computer system for storing and sorting the databases provided herein and practicing the methods  
10       provided herein.

Figure 18 is a flow diagram that illustrates the processing steps performed using the computer illustrated in Figure 17, to maintain and provide access to the databases for identifying polymorphic genetic markers.

Figure 19 is a histogram showing the allele and genotype distribution in  
15       the age and sex stratified Caucasian population for the AKAP10-1 locus. Bright green bars show frequencies in individuals younger than 40 years. Dark green bars show frequencies in individuals older than 60 years.

Figure 20 is a histogram showing the allele and genotype distribution in the age and sex stratified Caucasian population for the AKAP10-5 locus. Bright  
20       green bars show frequencies in individuals younger than 40 years; dark green bars show frequencies in individuals older than 60 years.

Figure 21 is a histogram showing the allele and genotype distribution in the age and sex stratified Caucasian population for the h-msrA locus. Genotype difference between male age groups is significant. Bright green bars show  
25       frequencies in individuals younger than 40 years. Dark green bars show frequencies in individuals older than 60 years.

Figure 22A-D is a sample data collection questionnaire used for the healthy database.

Figure 23 is a flowchart showing processing performed by the computing  
30       device of Figure 24 when performing genotyping of sense strands and antisense strands from assay fragments.

Figure 24 is a block diagram showing a system in accordance with the present invention;

Figure 25 is a flowchart of a method of identifying a biological sample in accordance with the present invention;

5        Figure 26 is a graphical representation of data from a mass spectrometer;

Figure 27 is a diagram of wavelet transformation of mass spectrometry data;

Figure 28 is a graphical representation of wavelet stage 0 hi data;

Figure 29 is a graphical representation of stage 0 noise profile;

10       Figure 30 is a graphical representation of generating stage noise standard deviations;

Figure 31 is a graphical representation of applying a threshold to data stages;

Figure 32 is a graphical representation of a sparse data set;

15       Figure 33 is a formula for signal shifting;

Figure 34 is a graphical representation of a wavelet transformation of a denoised and shifted signal;

Figure 35 is a graphical representation of a denoised and shifted signal;

Figure 36 is a graphical representation of removing peak sections;

20       Figure 37 is a graphical representation of generating a peak free signal ;

Figure 38 is a block diagram of a method of generating a baseline correction;

Figure 39 is a graphical representation of a baseline and signal;

Figure 40 is a graphical representation of a signal with baseline removed;

25       Figure 41 is a table showing compressed data;

Figure 42 is a flowchart of method for compressing data;

Figure 43 is a graphical representation of mass shifting;

Figure 44 is a graphical representation of determining peak width;

Figure 45 is a graphical representation of removing peaks;

30       Figure 46 is a graphical representation of a signal with peaks removed;

Figure 47 is a graphical representation of a residual baseline;

-10-

Figure 48 is a graphical representation of a signal with residual baseline removed;

Figure 49 is a graphical representation of determining peak height;

Figure 50 is a graphical representation of determining signal-to-noise for  
5 each peak;

Figure 51 is a graphical representation of determining a residual error for each peak;

Figure 52 is a graphical representation of peak probabilities;

Figure 53 is a graphical representation of applying an allelic ratio to peak  
10 probability;

Figure 54 is a graphical representation of determining peak probability

Figure 55 is a graphical representation of calling a genotype;

Figure 56 is a flowchart showing a statistical procedure for calling a  
genotype;

15 Figure 57 is a flowchart showing processing performed by the computing device of Figure 1 when performing standardless genotyping; and

Figure 58 is graphical representation of applying an allelic ratio to peak probability for standardless genotype processing.

#### DETAILED DESCRIPTION

##### 20 Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art to which this invention belongs. All patents, applications, published applications and other publications and sequences from GenBank and other  
25 databases referred to herein throughout the disclosure are incorporated by reference in their entirety.

As used herein, a biopolymer includes, but is not limited to, nucleic acid, proteins, polysaccharides, lipids and other macromolecules. Nucleic acids include DNA, RNA, and fragments thereof. Nucleic acids may be derived from  
30 genomic DNA, RNA, mitochondrial nucleic acid, chloroplast nucleic acid and other organelles with separate genetic material.

-11-

As used herein, morbidity refers to conditions, such as diseases or disorders, that compromise the health and well-being of an organism, such as an animal. Morbidity susceptibility or morbidity-associated genes are genes that, when altered, for example, by a variation in nucleotide sequence, facilitate the expression of a specific disease clinical phenotype. Thus, morbidity susceptibility genes have the potential, upon alteration, of increasing the likelihood or general risk that an organism will develop a specific disease.

As used herein, mortality refers to the statistical likelihood that an organism, particularly an animal, will not survive a full predicted lifespan.

10 Hence, a trait or a marker, such as a polymorphism, associated with increased mortality is observed at a lower frequency in older than younger segments of a population.

As used herein, a polymorphism, e.g. genetic variation, refers to a variation in the sequence of a gene in the genome amongst a population, such as allelic variations and other variations that arise or are observed. Thus, a polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. These differences can occur in coding and non-coding portions of the genome, and can be manifested or detected as differences in nucleic acid sequences, gene expression, including, for example transcription, processing, translation, transport, protein processing, trafficking, DNA synthesis, expressed proteins, other gene products or products of biochemical pathways or in post-translational modifications and any other differences manifested amongst members of a population. A single nucleotide polymorphism (SNP) refers to a polymorphism that arises as the result of a single base change, such as an insertion, deletion or change in a base.

A polymorphic marker or site is the locus at which divergence occurs. Such site may be as small as one base pair (an SNP). Polymorphic markers include, but are not limited to, restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats and other repeating patterns, simple sequence repeats and insertional elements, such as Alu. Polymorphic forms also are manifested as different mendelian

-12-

alleles for a gene. Polymorphisms may be observed by differences in proteins, protein modifications, RNA expression modification, DNA and RNA methylation, regulatory factors that alter gene expression and DNA replication, and any other manifestation of alterations in genomic nucleic acid or organelle nucleic acids.

5 As used herein, a healthy population, refers to a population of organisms, including but are not limited to, animals, bacteria, viruses, parasites, plants, eubacteria, and others, that are disease free. The concept of disease-free is a function of the selected organism. For example, for mammals it refers to a subject not manifesting any disease state. Practically a healthy subject, when  
10 human, is defined as human donor who passes blood bank criteria to donate blood for eventual use in the general population. These criteria are as follows: free of detectable viral, bacterial, mycoplasma, and parasitic infections; not anemic; and then further selected based upon a questionnaire regarding history (see Figure 3). Thus, a healthy population represents an unbiased population of  
15 sufficient health to donate blood according to blood bank criteria, and not further selected for any disease state. Typically such individuals are not taking any medications. For plants, for example, it is a plant population that does not manifest diseases pathology associated with plants. For bacteria it is a bacterial population replicating without environmental stress, such as selective agents,  
20 heat and other pathogens.

As used herein, a healthy database (or healthy patient database) refers to a database of profiles of subjects that have not been pre-selected for any particular disease. Hence, the subjects that serve as the source of data for the database are selected, according to predetermined criteria, to be healthy. In  
25 contrast to other such databases that have been pre-selected for subjects with a particular disease or other characteristic, the subjects for the database provided herein are not so-selected. Also, if the subjects do manifest a disease or other condition, any polymorphism discovered or characterized should be related to an independent disease or condition. In a preferred embodiment, where the  
30 subjects are human, a healthy subject manifests no disease symptoms and meets criteria, such as those set by blood banks for blood donors.

-13-

Thus, the subjects for the database are a population of any organism, including, but are not limited to, animals, plants, bacteria, viruses, parasites and any other organism or entity that has nucleic acid. Among preferred subjects are mammals, preferably, although not necessarily, humans. Such a database can  
5 capture the diversity of the a population, thus providing for discovery of rare polymorphisms.

As used herein, a profile refers to information relating to, but not limited to and not necessarily including all of, age, sex, ethnicity, disease history, family history, phenotypic characteristics, such as height and weight and other relevant  
10 parameters. A sample collect information form is shown in Figure 22, which illustrates profile intent.

As used herein, a disease state is a condition or abnormality or disorder that may be inherited or result from environmental stresses, such as toxins, bacterial, fungal and viral infections.

15 As used herein, set of non-selected subjects means that the subjects have not been pre-selected to share a common disease or other characteristic. They can be selected to be healthy as defined herein.

As used herein, a phenotype refers to a set of parameters that includes any distinguishable trait of an organism. A phenotype can be physical traits and  
20 can be, in instances in which the subject is an animal, a mental trait, such as emotional traits. Some phenotypes can be determined by observation elicited by questionnaires (see, *e.g.*, Figures 3 and 22) or by referring to prior medical and other records. For purposes herein, a phenotype is a parameter around which the database can be sorted.

25 As used herein, a parameter is any input data that will serve as a basis for sorting the database. These parameters will include phenotypic traits, medical histories, family histories and any other such information elicited from a subject or observed about the subject. A parameter may describe the subject, some historical or current environmental or social influence experienced by the  
30 subject, or a condition or environmental influence on someone related to the subject. Paramaters include, but are not limited to, any of those described herein, and known to those of skill in the art.

-14-

As used herein, haplotype refers to two or polymorphism located on a single DNA strand. Hence, haplotyping refers to identification of two or more polymorphisms on a single DNA strand. Haplotypes can be indicative of a phenotype. For some disorders a single polymorphism may suffice to indicate a trait; for others a plurality (i.e., a haplotype) may be needed. Haplotyping can be performed by isolating nucleic acid and separating the strands. In addition, when using enzymes such as certain nucleases, that produce, different size fragments from each strand, strand separation is not needed for haplotyping.

As used herein, used herein, pattern with reference to a mass spectrum or mass spectrometric analyses, refers to a characteristic distribution and number of signals (such peaks or digital representations thereof).

As used herein, signal in the context of a mass spectrum and analysis thereof refers to the output data, which the number or relative number of molecules having a particular mass. Signals include "peaks" and digital representations thereof.

As used herein, adaptor, when used with reference to haplotyping use Fen ligase, refers to a nucleic acid that specifically hybridizes to a polymorphism of interest. An adaptor can be partially double-stranded. An adaptor complex is formed when an adaptor hybridizes to its target.

As used herein, a target nucleic acid refers to any nucleic acid of interest in a sample. It can contain one or more nucleotides.

As used herein, standardless analysis refers to a determination based upon an internal standard. For example, the frequency of a polymorphism can be determined herein by comparing signals within a single mass spectrum.

As used herein, amplifying refers to means for increasing the amount of a bipolymer, especially nucleic acids. Based on the 5' and 3' primers that are chosen, amplification also serves to restrict and define the region of the genome which is subject to analysis. Amplification can be by any means known to those skilled in the art, including use of the polymerase chain reaction (PCR) etc.

Amplification, e.g., PCR must be done quantitatively when the frequency of polymorphism is required to be determined.

-15-

As used herein, cleaving refers to non-specific and specific fragmentation of a biopolymer.

As used herein, multiplexing refers to the simultaneous detection of more than one polymorphism. Methods for performing multiplexed reactions, particularly in conjunction with mass spectrometry are known (see, *e.g.*, U.S. Patent Nos. 6,043,031, 5,547,835 and International PCT application No. WO 97/37041).

As used herein, reference to mass spectrometry encompasses any suitable mass spectrometric format known to those of skill in the art. Such formats include, but are not limited to, Matrix-Assisted Laser Desorption/Ionization, Time-of-Flight (MALDI-TOF), Electrospray (ES), IR-MALDI (see, *e.g.*, published International PCT application No.99/57318 and U.S. Patent No. 5,118,937), Ion Cyclotron Resonance (ICR), Fourier Transform and combinations thereof. MALDI, particular UV and IR, are among the preferred formats.

As used herein, mass spectrum refers to the presentation of data obtained from analyzing a biopolymer or fragment thereof by mass spectrometry either graphically or encoded numerically.

As used herein, a blood component is a component that is separated from blood and includes, but is not limited to red blood cells and platelets, blood clotting factors, plasma, enzymes, plasminogen, immunoglobulins. A cellular blood component is a component of blood, such as a red blood cell, that is a cell. A blood protein is a protein that is normally found in blood. Examples of such proteins are blood factors VII and VIII. Such proteins and components are well-known to those of skill in the art.

As used herein, plasma can be prepared by any method known to those of skill in the art. For example, it can be prepared by centrifuging blood at a force that pellets the red cells and forms an interface between the red cells and the buffy coat, which contains leukocytes, above which is the plasma. For example, typical platelet concentrates contain at least about 10% plasma.

Blood may be separated into its components, including, but not limited to, plasma, platelets and red blood cells by any method known to those of skill in the art. For example, blood can be centrifuged for a sufficient time and at a

-16-

sufficient acceleration to form a pellet containing the red blood cells. Leukocytes collect primarily at the interface of the pellet and supernatant in the buffy coat region. The supernatant, which contains plasma, platelets, and other blood components, may then be removed and centrifuged at a higher acceleration, whereby the platelets pellet.

As used herein, p53 is a cell cycle control protein that assesses DNA damage and acts as a transcription factor regulation gene which control cell growth, DNA repair and apoptosis. The p53 mutations have been found in a wide variety of different cancers, including all of the different types of leukemia, with varying frequency. The loss of normal p53 functions results in genomic instability and uncontrolled growth of the host cell.

As used herein, p21 is a cyclin-dependent kinase inhibitor, associated with G1 phase arrest of normal cells. Expression triggers apoptosis or programmed cell death and has been associated with Wilms' tumor, a pediatric kidney cancer.

As used herein, Factor VII is a serine protease involved the extrinsic blood coagulation cascade. This factor is activated by thrombin and works with tissue factor (Factor III) in the processing of Factor X to Factor Xa. Evidence has supported an association between polymorphisms in the gene and increase Factor VII activity which can result in an elevated risk of ischemic cardiovascular disease including myocardial infarction.

As used herein, a relational database stores information in a form representative of matrices, such as two-dimensional tables, including rows and columns of data, or higher dimensional matrices. For example, in one embodiment, the relational database has separate tables each with a parameter. The tables are linked with a record number, which also acts as an index. The database can be searched or sorted by using data in the tables and is stored in any suitable storage medium, such as floppy disk, CD rom disk, hard drive or other suitable medium.

As used herein, a bar codes refers any array of optically readable marks of any desired size and shape that are arranged in a reference context or frame of, preferably, although not necessarily, one or more columns and one or more

-17-

rows. For purposes herein, the bar code refers to any symbology, not necessary "bar" but may include dots, characters or any symbol or symbols.

As used herein, symbology refers to an identifier code or symbol, such as a bar code, that is linked to a sample. The index will reference each such

5 symbology. The symbology is any code known or designed by the user. The symbols are associated with information stored in the database. For example, each sample can be uniquely identified with an encoded symbology. The parameters, such as the answers to the questions and subsequent genotypic and other information obtained upon analysis of the samples is included in the

10 database and associated with the symbology. The database is stored on any suitable recording medium, such as a hard drive, a floppy disk, a tape, a CD ROM, a DVD disk and any other suitable medium.

#### DATABASES

Human genotyping is currently dependent on collaborations with

15 hospitals, tissues banks and research institutions that provide samples of disease tissue. This approach is based on the concept that the onset and/or progression of diseases can be correlated with the presence of a polymorphisms or other genetic markers. This approach does not consider that disease correlated with the presence of specific markers and the absence of specific markers. It is

20 shown herein that identification and scoring of the appearance and disappearance of markers is possible only if these markers are measured in the background of healthy subjects where the onset of disease does not mask the change in polymorphism occurrence. Databases of information from disease populations suffer from small sample size, selection bias and heterogeneity. The

25 databases provided herein from healthy populations solve these problems by permitting large sample bands, simple selection methods and diluted heterogeneity.

Provided herein are first databases of parameters, associated with non-selected, particularly healthy, subjects. Also provided are combinations of the

30 databases with indexed samples obtained from each of the subjects. Further provided are databases produced from the first databases. These contain in addition to the original parameters information, such as genotypic information,

-18-

including, but are not limited to, genomic sequence information, derived from the samples.

The databases, which are herein designated healthy databases, are so-designated because they are not obtained from subjects pre-selected for a particular disease. Hence, although individual members may have a disease, the collection of individuals is not selected to have a particular disease.

The subjects from whom the parameters are obtained comprise either a set of subjects who are randomly selected across, preferably, all populations, or are pre-selected to be disease-free or healthy. As a result, the database is not selected to be representative of any pre-selected phenotype, genotype, disease or other characteristic. Typically the number of subjects from which the database is prepared is selected to produce statistically significant results when used in the methods provided herein. Preferably, the number of subjects will be greater than 100, more preferably greater than 200, yet more preferably greater than 1000. The precise number can be empirically determined based upon the frequency of the parameter(s) that be used to sort the database. Generally the population can have at least 50, at least 100, at least 200, at least 500, at least 1000, at least 5000 or at least 10,000 or more subjects.

Upon identification of a collection of subjects, information about each subject is recorded and associated with each subject as a database. The information associated with each of the subjects, includes, but is not limited to, information related to historical characteristics of the subjects, phenotypic characteristics and also genotypic characteristics, medical characteristics and any other traits and characteristics about the subject that can be determined. This information will serve as the basis for sorting the database.

In an exemplary embodiment, the subjects are mammals, such as humans, and the information relates to one or more of parameters, such as age, sex, medical history, ethnicity and any other factor. Such information, when the animals are humans, for example, can be obtained by a questionnaire, and by observations about the individual, such as hair color, eye color and other characteristics. Genotypic information will be obtained from tissue or other body and body fluid samples from the subject.

-19-

The healthy genomic database can include profiles and polymorphisms from healthy individuals from a library of blood samples where each sample in the library is an individual and separate blood or other tissue sample. Each sample in the database is profiled as to the sex, age, ethnic group, and disease history of the donor.

The databases are generated by first identifying healthy populations of subjects and obtaining information about each subject that will serve as the sorting parameters for the database. This information is preferably entered into a storage medium, such as the memory of a computer.

The information obtained about each subject in a population used for generating the database is stored in a computer memory or other suitable storage medium. The information is linked to an identifier associated with each subject. Hence the database will identify a subject, for example by a datapoint representative of a bar code, and then all information, such as the information from a questionnaire, regarding the individual is associated with the datapoint. As the information is collected the database is generated.

Thus, for example, profile information, such as subject histories obtained from questionnaires, is collected in the database. The resulting database can be sorted as desired, using standard software, such as by age, sex and/or ethnicity.

An exemplary questionnaire for subjects from whom samples are to be obtained is shown in Figures 22A-D. Each questionnaire preferably is identified by a bar code, particularly a machine readable bar code for entry into the database. After a subject provides data and is deemed to be healthy (*i.e.*, meets standards for blood donation), the data in the questionnaire is entered into the database and is associated with the bar code. A tissue, cell or blood sample is obtained from the subject.

Figure 4 exemplifies processing and tracking of blood sample components. Each component is tracked with a bar code, dated, is entered into the database and associated with the subject and the profile of the subject.

Typically, the whole blood is centrifuged to produce plasma, red blood cells (which pellet) and leukocytes found in the buffy coat which layers in between.

-20-

Various samples are obtained and coded with a bar code and stored for use as needed.

Samples are collected from the subjects. The samples include, but are not limited to, tissues, cells, and fluids, such as nucleic acid, blood, plasma, amniotic fluid, synovial fluid, urine, saliva, aqueous humor, sweat, sperm samples and cerebral spinal fluid. It is understood that the particular set of samples depends upon the organisms in the population.

Once samples are obtained the collection can be stored and, in preferred embodiments, each sample is indexed with an identifier, particularly a machine readable code, such as a bar code. For analyses, the samples or components of the samples, particularly biopolymers and small molecules, such as nucleic acids and/or proteins and metabolites, are isolated.

After samples are analyzed, this information is entered into the database in the memory of the storage medium and associated with each subject. This information includes, but is not limited to, genotypic information. Particularly, nucleic acid sequence information and other information indicative of polymorphisms, such as masses of PCR fragments, peptide fragment sequences or masses, spectra of biopolymers and small molecules and other indicia of the structure or function of a gene, gene product or other marker from which the existence of a polymorphism within the population can be inferred.

In an exemplary embodiment, a database can be derived from a collection of blood samples. For example, Figure 1 (see, also Figure 10) shows the status of a collection of over 5000 individual samples. The samples were processed in the laboratory following SOP (standard operating procedure) guidelines. Any standard blood processing protocol may be used.

For the exemplary database described herein, the following criteria were used to select subjects:

No testing is done for infectious agents.

Age: At least 17 years old

Weight: Minimum of 110 pounds

Permanently Disqualified:

History of hepatitis (after age 11)

-21-

Leukemia Lymphoma

Human immunodeficiency virus (HIV), AIDS

Chronic kidney disease

Temporarily Disqualified:

- 5       Pregnancy - until six weeks after delivery, miscarriage or abortion  
Major surgery or transfusions - for one year  
Mononucleosis - until complete recovery  
Prior whole blood donation - for eight weeks  
Antibiotics by injection for one week; by mouth, for forty-eight hours,  
10       except antibiotics for skin complexion;

5 year Deferment:

Internal cancer and skin cancer if it has been removed, is healed and  
there is no recurrence

These correspond to blood bank criteria for donating blood and represent a

- 15       healthy population as defined herein for a human healthy database.

**Structure of the database**

- Any suitable database structure and format known to those of skill in the  
art may be employed. For example, a relational database is a preferred format in  
which data is stored as matrices or tables of the parameters linked by an indexer  
20       that identifies each subject. Software for preparing and manipulating, including  
sorting the database, can be readily developed or adapted from commercially  
available software, such as Microsoft Access.

**Quality control**

- Quality control procedures can be implemented. For example, after  
25       collection of samples, the quality of the collection in the bank can be assessed.  
For example, mix-up of samples can be checked by testing for known markers,  
such as sex. After samples are separated by ethnicity, samples are randomly  
tested for a marker associated with a particular ethnicity, such as HLA DQA1  
group specific component, to assess whether the samples have been properly  
30       sorted by ethnic group. An exemplary sample bank is depicted in Figure 4.

-22-

**Obtaining genotypic data and other parameters for the database**

After informational and historical parameters are entered into the database, material from samples obtained from each subject, is analyzed. Analyzed material include proteins, metabolites, nucleic acids, lipids and any  
5 other desired constituent of the material. For example, nucleic acids, such as genomic DNA, can be analyzed by sequencing.

Sequencing can be performed using any method known to those of skill in the art. For example, if a polymorphism is identified or known, and it is desired to assess its frequency or presence among the subjects in the database, the  
10 region of interest from each sample can be isolated, such as by PCR or restriction fragments, hybridization or other suitable method known to those of skill in the art and sequenced. For purposes herein, sequencing analysis is preferably effected using mass spectrometry (see, *e.g.*, U.S. Patent Nos. 5,547,835, 5,622,824, 5,851,765, and 5,928,906). Nucleic acids can also be  
15 sequence by hybridization (see, *e.g.*, U.S. Patent Nos. 5,503,980, 5,631,134, 5,795,714) and including analysis by mass spectrometry (see, U.S. application Serial Nos. 08/419,994 and 09/395,409).

In other detection methods, it is necessary to first amplify prior to identifying the allelic variant. Amplification can be performed, *e.g.*, by PCR  
20 and/or LCR, according to methods known in the art. In one embodiment, genomic DNA of a cell is exposed to two PCR primers and amplification for a number of cycles sufficient to produce the required amount of amplified DNA. In preferred embodiments, the primers are located between 150 and 350 base pairs apart.

25 Alternative amplification methods include: self sustained sequence replication (Guatelli, J. C. et al., 1990, Proc. Natl. Acad. Sci. U.S.A. 87:1874-1878), transcriptional amplification system (Kwoh, D. Y. et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:1173-1177), Q-Beta Replicase (Lizardi, P. M. et al., 1988, Bio/Technology 6:1197), or any other nucleic acid amplification  
30 method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially

-23-

useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

- Nucleic acids can also be analyzed by detection methods and protocols, particularly those that rely on mass spectrometry (see, *e.g.*, U.S. Patent No. 5,605,798, 6,043,031, allowed copending U.S. application Serial No. 08/744,481, U.S. application Serial No. 08/990,851 and International PCT application No. WO 99/31273, International PCT application No. WO 98/20019). These methods can be automated (see, *e.g.*, copending U.S. application Serial No. 09/285,481 and published International PCT application No. PCT/US00/08111, which describes an automated process line). Preferred among the methods of analysis herein are those involving the primer oligo base extension (PROBE) reaction with mass spectrometry for detection (described herein and elsewhere, see *e.g.*, U.S. Patent No. 6,043,031; see, also U.S. application Serial Nos. 09/287,681, 09/287,682, 09/287,141 and 09/287,679, allowed copending U.S. application Serial No. 08/744,481, International PCT application No. PCT/US97/20444, published as International PCT application No. WO 98/20019, and based upon U.S. application Serial Nos. 08/744,481, 08/744,590, 08/746,036, 08/746,055, 08/786,988, 08/787,639, 08/933,792, 08/746,055, 08/786,988 and 08/787,639; see, also U.S. application Serial No. 09/074,936, U.S. Patent No. 6,024,925, and U.S. application Serial Nos. 08/746,055 and 08/786,988, and published International PCT application No. WO 98/20020)

- A preferred format for performing the analyses is a chip based format in which the biopolymer is linked to a solid support, such as a silicon or silicon-coated substrate, preferably in the form of an array. More preferably, when analyses are performed using mass spectrometry, particularly MALDI, small nanoliter volumes of sample are loaded on, such that the resulting spot is about, or smaller than, the size of the laser spot. It has been found that when this is achieved, the results from the mass spectrometric analysis are quantitative. The area under the signals in the resulting mass spectra are proportional to concentration (when normalized and corrected for background). Methods for preparing and using such chips are described in U.S. Patent No. 6,024,925, co-

-24-

pending U.S. application Serial Nos. 08/786,988, 09/364,774, 09/371,150 and 09/297,575; see, also U.S. application Serial No. PCT/US97/20195, which published as WO 98/20020. Chips and kits for performing these analyses are commercially available from SEQUENOM under the trademark MassARRAY.

5 MassArray relies on the fidelity of the enzymatic primer extension reactions combined with the miniaturized array and MALDI-TOF (Matrix-Assisted Laser Desorption Ionization-Time of Flight) mass spectrometry to deliver results rapidly. It accurately distinguishes single base changes in the size of DNA fragments associated with genetic variants without tags.

10 The methods provided herein permit quantitative determination of alleles. The areas under the signals in the mass spectra can be used for quantitative determinations. The frequency is determined from the ratio of the signal to the total area of all of the spectrum and corrected for background. This is possible because of the PROBE technology as described in the above applications

15 incorporated by reference herein.

Additional methods of analyzing nucleic acids include amplification- based methods including polymerase chain reaction (PCR), ligase chain reaction (LCR), mini-PCR, rolling circle amplification, autocatalytic methods, such as those using Q $\beta$  replicase, TAS, 3SR, and any other suitable method known to those of skill

20 in the art.

Other methods for analysis and identification and detection of polymorphisms, include but are not limited to, allele specific probes, Southern analyses, and other such analyses.

The methods described below provide ways to fragment given amplified

25 or non-amplified nucleotide sequences thereby producing a set of mass signals when mass spectrometry is used to analyze the fragment mixtures.

Amplified fragments are yielded by standard polymerase chain methods (US 4,683,195 and 4,683,202). The fragmentation method involves the use of enzymes that cleave single or double strands of DNA and enzymes that ligate

30 DNA. The cleavage enzymes can be glycosylases, nickases, and site-specific and non site-specific nucleases with the most preferred enzymes being glycosylases, nickases, and site-specific nucleases.

-25-

### Glycosylase Fragmentation Method

DNA glycosylases specifically remove a certain type of nucleobase from a given DNA fragment. These enzymes can thereby produce abasic sites, which can be recognized either by another cleavage enzyme, cleaving the exposed phosphate backbone specifically at the abasic site and producing a set of nucleobase specific fragments indicative of the sequence, or by chemical means, such as alkaline solutions and or heat. The use of one combination of a DNA glycosylase and its targeted nucleotide would be sufficient to generate a base specific signature pattern of any given target region.

- 10 Numerous DNA glycosylases are known. For example, a DNA glycosylase can be uracil-DNA glycosylase (UDG), 3-methyladenine DNA glycosylase, 3-methyladenine DNA glycosylase II, pyrimidine hydrate-DNA glycosylase, FaPy-DNA glycosylase, thymine mismatch-DNA glycosylase, hypoxanthine-DNA glycosylase, 5-Hydroxymethyluracil DNA glycosylase (HmUDG), 5-Hydroxymethylcytosine DNA glycosylase, or 1,N6-ethenoadenine DNA glycosylase (see, *e.g.*, U.S. Patent Nos. 5,536,649, 5,888, 795, 5,952,176 and 6,099,553, International PCT application Nos. WO 97/03210, WO 99/54501; see, also, Eftedal et al. (1993) *Nucleic Acids Res* 21:2095-2101, Bjelland and Seeberg (1987) *Nucleic Acids Res.* 15:2787-2801, Saparbaev et al. (1995) *Nucleic Acids Res.* 23:3750-3755, Bessho (1999) *Nucleic Acids Res.* 27:979-983) corresponding to the enzyme's modified nucleotide or nucleotide analog target. A preferred glycosylase is uracil-DNA glycosylase (UDG).

- 25 Uracil, for example, can be incorporated into an amplified DNA molecule by amplifying the DNA in the presence of normal DNA precursor nucleotides (e.g. dCTP, dATP, and dGTP) and dUTP. When the amplified product is treated with UDG, uracil residues are cleaved. Subsequent chemical treatment of the products from the UDG reaction results in the cleavage of the phosphate backbone and the generation of nucleobase specific fragments. Moreover, the separation of the complementary strands of the amplified product prior to glycosylase treatment allows complementary patterns of fragmentation to be generated. Thus, the use of dUTP and Uracil DNA glycosylase allows the generation of T specific fragments for the complementary strands, thus providing

-26-

information on the T as well as the A positions within a given sequence. Similar to this, a C-specific reaction on both (complementary) strands (i.e. with a C-specific glycosylase) yields information on C as well as G positions within a given sequence if the fragmentation patterns of both amplification strands are  
5 analyzed separately. Thus, with the glycosylase method and mass spectrometry, a full series of A, C, G and T specific fragmentation patterns can be analyzed.

#### Nickase Fragmentation Method

A DNA nickase, or DNase, can be used recognize and cleave one strand  
10 of a DNA duplex. Numerous nickases are known. Among these, for example, are nickase NY2A nickase and NYS1 nickase (Megabase) with the following cleavage sites:

NY2A: 5'...R AG...3'

3'...Y TC...5' where R = A or G and Y = C or T

15 NYS1: 5'... CC[A/G/T]...3'

3'... GG[T/C/A]...5'.

#### Fen-Ligase Fragmentation Method

The Fen-ligase method involves two enzymes: Fen-1 enzyme and a ligase. The Fen-1 enzyme is a site-specific nuclease known as a "flap" endonuclease  
20 (US 5,843,669, 5,874,283, and 6,090,606). This enzymes recognizes and cleaves DNA "flaps" created by the overlap of two oligonucleotides hybridized to a target DNA strand. This cleavage is highly specific and can recognize single base pair mutations, permitting detection of a single homologue from an individual heterozygous at one SNP of interest and then genotyping that  
25 homologue at other SNPs occurring within the fragment. Fen-1 enzymes can be Fen-1 like nucleases e.g. human, murine, and Xenopus XPG enzymes and yeast RAD2 nucleases or Fen-1 endonucleases from, for example, *M. jannaschii*, *P. furiosus*, and *P. woesei*. Among preferred enzymes are the Fen-1 enzymes.

The ligase enzyme forms a phosphodiester bond between two double  
30 stranded nucleic acid fragments. The ligase can be DNA Ligase I or DNA Ligase III (see, e.g., U.S. Patent Nos. US 5,506,137, 5,700,672, 5,858,705 and 5,976,806; see, also, Waga, *et al.* (1994) J. Biol. Chem. 269:10923-10934, Li

-27-

et al. (1994) Nucleic Acids Res. 22:632-638, Arrand et al. (1986) J. Biol. Chem. 261:9079-9082, Lehman (1974) Science 186:790-797, Higgins and Cozzarelli (1979) Methods Enzymol. 68:50-71, Lasko et al. (1990) Mutation Res. 236:277-287, and Lindahl and Barnes (1992) Ann. Rev. Biochem. 61:251-281).

- 5           Thermostable ligase (Epicenter Technologies), where "thermostable" denotes that the ligase retains activity even after exposure to temperatures necessary to separate two strands of DNA, are among preferred ligases for use herein.

#### Type IIS Enzyme Fragmentation Method

- 10           Restriction enzymes bind specifically to and cleave double-stranded DNA at specific sites within or adjacent to a particular recognition sequence. These enzymes have been classified into three groups (e.g. Types I, II, and III) as known to those of skill in the art. Because of the properties of type I and type III enzymes, they have not been widely used in molecular biological applications.
- 15           Thus, for this invention type II enzymes are preferred. Of the thousands of restriction enzymes known in the arts, there are 179 different type II specificities. Of the 179 unique type II restriction endonucleases, 31 have a 4-base recognition sequence, 11 have a 5-base recognition sequence, 127 have a 6-base recognition sequence, and 10 have recognition sequences of greater than
- 20           six bases (US 5,604,098). Of category type II enzymes, type IIS is preferred.

          Type IIS enzymes can be *Alw* XI, *Bbv* I, *Bce* 83, *Bpm* I, *Bsg* I, *Bsm* AI, *Bsm* FI, *Bsa* I, *Bcc* I, *Bcg* I, *Ear* I, *Eco* 57I, *Esp* 3I, *Fau* I, *Fok* I, *Gsu* I, *Hga* I, *Mme* I, *Mbo* II, *Sap* I, and the like. The preferred type IIS enzyme is *Fok* I.

- The *Fok* I enzyme endonuclease is an exemplary well characterized
- 25           member of the Type IIS class (see, e.g., U.S. Patent Nos. 5,714,330, 5,604,098, 5,436,150, 6,054,276 and 5,871,911; see, also, Szybalski et al. (1991) Gene 100:13-26, Wilson and Murray (1991) Ann. Rev. Genet. 25:585-627, Sugisaki et al. (1981) Gene 16:73-78, Podhajska and Szalski (1985) Gene 40:175-182. *Fok* I recognizes the sequence 5'GGATG-3' and cleaves DNA
- 30           accordingly. Type IIS restriction sites can be introduced into DNA targets by incorporating the site into primers used to amplify such targets. Fragments produced by digestion with *Fok* I are site specific and can be analyzed by mass

-28-

spectrometry methods such as MALDI-TOF mass spectrometry, ESI-TOF mass spectrometry, and any other type of mass spectrometry well known to those of skill in the art.

- Once a polymorphism has been found to correlatate with a parameter such as age. The possibility of false results due to allelic dropout is examined by doing comparative PCR in an adjacent region of the genome.

#### Analyses

- In using the database, allelic frequencies can be determined across the population by analyzing each sample in the population individually, determining the presence or absence of allele or marker of interest in each individual sample, and then determining the frequency of the marker in the population. The database can then be sorted (stratified) to identify any correlations between the allele and a selected parameter using standard statistical analysis. If a correlation is observed, such as a decrease in a particular marker with age or correlation with sex or other parameter, then the marker is a candidate for further study, such as genetic mapping to identify a gene or pathway in which it is involved. The marker may then be correlated, for example, with a disease. Haplotyping can also be carried out. Genetic mapping can be effected using standard methods and may also require use of databases of others, such as databases previously determined to be associated with a disorder.

Exemplary analyses have been performed and these are shown in the figures, and discussed herein.

#### Sample pooling

- It has been found that using the databases provided herein, or any other database of such information, substantially the same frequencies that were obtained by examining each sample separately can be obtained by pooling samples, such as in batches of 10, 20, 50, 100, 200, 500, 1000 or any other number. A precise number may be determined empirically if necessary, and can be as low as 3.

-29-

In one embodiment, the frequency of genotypic and other markers can be obtained by pooling samples. To do this a target population and a genetic variation to be assessed is selected, a plurality of samples of biopolymers are obtained from members of the population, and the biopolymer from which the  
5 marker or genotype can be inferred is determined or detected. A comparison of samples tested in pools and individually and the sorted results therefrom are shown in Figure 9, which shows frequency of the factor VII Allele 353Q. Figure 10 depicts the frequency of the CETP Allele CETP in pooled versus individual samples. Figure 15 shows ethnic diversity among various ethnic groups in the  
10 database using pooled DNA samples to obtain the data. Figures 12-14 show mass spectra for these samples.

Pooling of test samples has application not only to the healthy databases provided herein, but also to use in gathering data for entry into any database of subjects and genotypic information, including typical databases derived from  
15 diseased populations. What is demonstrated herein, is the finding that the results achieved are statistically the same as the results that would be achieved if each sample is analyzed separately. Analysis of pooled samples by a method, such as the mass spectrometric methods provided herein, permits resolution of such data and quantitation of the results.

20 For factor VII the R53Q acid polymorphism was assessed. In Figure 9, the "individual" data represent allelic frequency observed in 92 individuals reactions. The pooled data represent the allelic frequency of the same 92 individuals pooled into a single probe reaction. The concentration of DNA in the samples of individual donors is 250 nanograms. The total concentration of DNA  
25 in the pooled samples is also 250 nanograms, where the concentration of any individual DNA is 2.7 nanograms.

It also was shown that it is possible to reduce the DNA concentration of individuals in a pooled samples from 2.7 nanograms to 0.27 nanograms without any change in the quality of the spectrum or the ability to quantitate the amount  
30 of sample detected. Hence low concentrations of sample may be used in the pooling methods.

-30-

**Use of the databases and markers identified thereby**

The successful use of genomics requires a scientific hypothesis (*i.e.*, common genetic variation, such as a SNP), a study design (*i.e.*, complex disorders), samples and technology, such as the chip-based mass spectrometric analyses (see, *e.g.*, U.S. Patent No. 5,605,798, U.S. Patent No. 5,777,324, U.S. Patent No. 6,043,031, allowed copending U.S. application Serial No. 08/744,481, U.S. application Serial No. 08/990,851, International PCT application No. WO 98/20019, copending U.S. application Serial No. 09/285,481, which describes an automated process line for analyses; see, also, 10 U.S. application Serial Nos. 08/617,256, 09/287,681, 09/287,682, 09/287,141 and 09/287,679, allowed copending U.S. application Serial No. 08/744,481, International PCT application No. PCT/US97/20444, published as International PCT application No. WO 98/20019, and based upon U.S. application Serial Nos. 08/744,481, 08/744,590, 08/746,036, 08/746,055, 08/786,988, 08/787,639, 15 08/933,792, 08/746,055, 09/266,409, 08/786,988 and 08/787,639; see, also U.S. application Serial No. 09/074,936). All of these aspects can be used in conjunction with the databases provided herein and samples in the collection.

The databases and markers identified thereby can be used, for example, for identification of previously unidentified or unknown genetic markers and to 20 identify new uses for known markers. As markers are identified, these may be entered into the database to use as sorting parameters from which additional correlations may be determined.

**Previously unidentified or unknown genetic markers**

The samples in the healthy databases can be used to identify new 25 polymorphisms and genetic markers, using any mapping, sequencing, amplification and other methodologies, and in looking for polymorphisms among the population in the database. The thus-identified polymorphism can then be entered into the database for each sample, and the database sorted (stratified) using that polymorphism as a sorting parameter to identify any patterns and 30 correlations that emerge, such as age correlated changes in the frequency of the identified marker. If a correlation is identified, the locus of the marker can be mapped and its function or effect assessed or deduced.

-31-

Thus, the databases here provide means for:

identification of significantly different allelic frequencies of genetic factors by comparing the occurrence or disappearance of the markers with increasing age in population and then associating the markers with a disease or a

5 biochemical pathway;

identification of significantly different allelic frequencies of disease causing genetic factors by comparing the male with the female population or comparing other selected stratified populations and associating the markers with a disease or a biochemical pathway;

10 identification of significantly different allelic frequencies of disease causing genetic factors by comparing different ethnic groups and associating the markers with a disease or a biochemical pathway that is known to occur in high frequency in the ethnic group;

profiling potentially functional variants of genes through the general  
15 panmixed population stratified according to age, sex, and ethnic origin and thereby demonstrating the contribution of the variant genes to the physical condition of the investigated population;

identification of functionally relevant gene variants by gene disequilibrium analysis performed within the general panmixed population stratified according  
20 to age, sex, and ethnic origin and thereby demonstrating their contribution to the physical condition of investigated population;

identification of potentially functional variants of chromosomes or parts of chromosomes by linkage disequilibrium analysis performed within the general panmixed population stratified according to age, sex, and ethnic origin and  
25 thereby demonstrating their contribution to the physical condition of investigated population.

#### **Uses of the identified markers and known markers**

The databases may also be used in conjunction with known markers and sorted to identify any correlations. For example, the databases can be used for:

30 determination and evaluation of the penetrance of medically relevant polymorphic markers;

-32-

determination and evaluation of the diagnostic specificity of medically relevant genetic factors;

determination and evaluation of the positive predictive value of medically relevant genetic factors;

- 5           determination and evaluation of the onset of complex diseases, such as, but are not limited to, diabetes, hypertension, autoimmune diseases, arteriosclerosis, cancer and other diseases within the general population with respect to their causative genetic factors;

delineation of the appropriate strategies for preventive disease treatment;

- 10          delineation of appropriate timelines for primary disease intervention;

validation of medically relevant genetic factors identified in isolated populations regarding their general applicability;

validation of disease pathways including all potential target structures identified in isolated populations regarding their general applicability; and

- 15          validation of appropriate drug targets identified in isolated populations regarding their general applicability.

Among the diseases and disorders for which polymorphisms may be linked include, those linked to inborn errors of metabolism, acquired metabolic disorders, intermediary metabolism, oncogenesis pathways, blood clotting

- 20          pathways, and DNA synthetic and repair pathways DNA repair/replication/transcription factors and activities, *e.g.*, such as genes related to oncogenesis, aging and genes involved in blood clotting and the related biochemical pathways that are related to thrombosis, embolism, stroke, myocardial infarction, angiogenesis and oncogenesis.

- 25          For example, a number of diseases are caused by or involve deficient or defective enzymes in intermediary metabolism (see, e.g., Tables 1 and 2, below) that result, upon ingestion of the enzyme substrates, in accumulation of harmful metabolites that damage organs and tissues, particularly an infant's developing brain and other organs, resulting in mental retardation and other developmental
- 30          disorders.

-33-

Identification of markers and genes for such disorders is of great interest.

#### Model systems

Several gene systems, p21, p53 and Lipoprotein Lipase polymorphism (N291S), were selected. The p53 gene is a tumor suppressor gene that is  
5 mutated in diverse tumor types. One common allelic variant occurs at codon 72. A polymorphism that has been identified in the p53 gene, i.e., the R72P allele, results in an amino acid exchange, arginine to proline, at codon 72 of the gene.

Using diseased populations, it has been shown that there are ethnic differences in the allelic distribution of these alleles among African-Americans  
10 and Caucasians in the U.S. The results here support this finding and also demonstrate that the results obtained with a healthy database are meaningful (see, Figure 7B).

The 291S allele leads to reduced levels of high density lipoprotein cholesterol (HDL-C) that is associated with an increased risk of males for  
15 arteriosclerosis and in particular myocardial infarction (see, Reymer *et al.* (1995) *Nature Genetics* 10:28-34).

Both genetic polymorphisms were profiled within a part of the Caucasian population-based sample bank. For the polymorphism located in the lipoprotein lipase gene a total of 1025 unselected individuals (436 males and 589 females)  
20 were tested. Genomic DNA was isolated from blood samples obtained from the individuals.

As shown in the Examples and figures, an exemplary database containing about 5000 subjects, answers to the questionnaire (see Figure 3), and genotypic information has been stratified. A particular known allele has been selected, and  
25 the samples tested for the marker using mass spectrometric analyses, particularly PROBE (see the EXAMPLES) to identify polymorphisms in each sample. The population in the database has been sorted according to various parameters and correlations have been observed. For example, FIGURES 2A-C, show sorting of the data by age and sex for the Lipoprotein Lipase gene in the  
30 Caucasian population in the database. The results show a decrease in the frequency of the allele with age in males but no such decrease in females. Other

-34-

alleles that have been tested against the database, include, alleles of p53, p21 and factor VII. Results when sorted by age are shown in the figures.

These examples demonstrate an effect of altered frequency of disease causing genetic factors within the general population. The scientific

5 interpretation of those results allows prediction of medical relevance of polymorphic genetic alterations. In addition, conclusions can be drawn with regard to their penetrance, diagnostic specificity, positive predictive value, onset of disease, most appropriate onset of preventive strategies, and the general applicability of genetic alterations identified in isolated populations to panmixed  
10 populations.

Therefore, an age- and sex-stratified population-based sample bank that is ethnically homogenous is a suitable tool for rapid identification and validation of genetic factors regarding their potential medical utility.

15 **Exemplary computer system for creating, storing and processing the databases Systems**

Systems, including computers, containing the databases are provided herein. The computers and databases can be used in conjunction, for example, with the APL system (see, copending U.S. application Serial No. 09/285,481),  
20 which is an automated system for analyzing biopolymers, particularly nucleic acids. Results from the APL system can be entered into the database.

Any suitable computer system may be used. The computer system may be integrated into systems for sample analysis, such as the automated process line described herein (see, *e.g.*, copending U.S. application Serial No.  
25 09/285,481).

Figure 17 is a block diagram of a computer constructed in to provide and process the databases described herein. The processing that maintains the database and performs the methods and procedures may be performed on multiple computers all having a similar construction, or may be performed by a  
30 single, integrated computer. For example, the computer through which data is added to the database may be separate from the computer through which the database is sorted, or may be integrated with it. In either arrangement, the

-35-

computers performing the processing may have a construction as illustrated in Figure 17.

Figure 17 is a block diagram of an exemplary computer 1700 that maintains the database described above and performs the methods and

5 procedures. Each computer 1700 operates under control of a central processor unit (CPU) 1702, such as a "Pentium" microprocessor and associated integrated circuit chips, available from Intel Corporation of Santa Clara, California, USA. A computer user can input commands and data from a keyboard and display

10 mouse 1704 and can view inputs and computer output at a display 1706. The display is typically a video monitor or flat panel display device. The computer 1700 also includes a direct access storage device (DASD) 1707, such as a fixed hard disk drive. The memory 1708 typically comprises volatile semiconductor random access memory (RAM). Each computer preferably includes a program

15 product reader 1710 that accepts a program product storage device 1712, from which the program product reader can read data (and to which it can optionally write data). The program product reader can comprise, for example, a disk drive, and the program product storage device can comprise removable storage media such as a magnetic floppy disk, an optical CD-ROM disc, a CD-R disc, a CD-RW disc, or a DVD data disc. If desired, the computers can be connected so

20 they can communicate with each other, and with other connected computers, over a network 1713. Each computer 1700 can communicate with the other connected computers over the network 1713 through a network interface 1714 that enables communication over a connection 1716 between the network and the computer.

25 The computer 1700 operates under control of programming steps that are temporarily stored in the memory 1708 in accordance with conventional computer construction. When the programming steps are executed by the CPU 1702, the pertinent system components perform their respective functions. Thus, the programming steps implement the functionality of the system as

30 described above. The programming steps can be received from the DASD 1707, through the program product reader 1712, or through the network connection 1716. The storage drive 1710 can receive a program product, read

-36-

programming steps recorded thereon and transfer the programming steps into the memory 1708 for execution by the CPU 1702. As noted above, the program product storage device 1710 can comprise any one of multiple removable media having recorded computer-readable instructions, including

5 magnetic floppy disks and CD-ROM storage discs. Other suitable program product storage devices can include magnetic tape and semiconductor memory chips. In this way, the processing steps necessary for operation can be embodied on a program product.

Alternatively, the program steps can be received into the operating

10 memory 1708 over the network 1713. In the network method, the computer receives data including program steps into the memory 1708 through the network interface 1714 after network communication has been established over the network connection 1716 by well-known methods that will be understood by those skilled in the art without further explanation. The program steps are then

15 executed by the CPU 1702 to implement the processing of the Garment Database system.

It should be understood that all of the computers of the system preferably have a construction similar to that shown in Figure 17, so that details described with respect to the Figure 17 computer 1700 will be understood to apply to all

20 computers of the system 1700. This is indicated by multiple computers 1700 shown connected to the network 1713. Any one of the computers 1700 can have an alternative construction, so long as they can communicate with the other computers and support the functionality described herein.

Figure 18 is a flow diagram that illustrates the processing steps

25 performed using the computer illustrated in Figure 17, to maintain and provide access to the databases, such as for identifying polymorphic genetic markers. In particular, the information contained in the database is stored in computers having a construction similar to that illustrated in Figure 17. The first step for maintaining the database, as indicated in Figure 18, is to identify healthy

30 members of a population. As noted above, the population members are subjects that are selected only on the basis of being healthy, and where the subjects are mammals, such as humans, they are preferably selected based upon apparent

-37-

health and the absence of detectable infections. The step of identifying is represented by the flow diagram box numbered 1802.

The next step, represented by the flow diagram box numbered 1804, is to obtain identifying and historical information and data relating to the identified members of the population. The information and data comprise parameters for each of the population members, such as member age, ethnicity, sex, medical history, and ultimately genotypic information. Initially, the parameter information is obtained from a questionnaire answered by each member, from whom a body tissue or body fluid sample also is obtained. The step of entering and storing these parameters into the database of the computer is represented by the flow diagram box numbered 1806. As additional information about each population member and corresponding sample is obtained, this information can be inputted into the database and can serve as a sorting parameter.

In the next step, represented by the flow diagram box numbered 1808, the parameters of the members are associated with an indexer. This step may be executed as part of the database storage operation, such as when a new data record is stored according to the relational database structure and is automatically linked with other records according to that structure. The step 1806 also may be executed as part of a conventional data sorting or retrieval process, in which the database entries are searched according to an input search or indexing key value to determine attributes of the data. For example, such search and sort techniques may be used to follow the occurrence of known genetic markers and then determine if there is a correlation with diseases for which they have been implicated. Examples of this use are for assessing the frequencies of the p53 and Lipoprotein Lipase polymorphisms.

Such searching of the database also may be valuable for identifying one or more genetic markers whose frequency changes within the population as a function of age, ethnic group, sex, or some other criteria. This can allow the identification of previously unknown polymorphisms and, ultimately, identification of a gene or pathway involved in the onset and progression of disease.

-38-

In addition, the database can be used for taking an identified polymorphism and ascertaining whether it changes in frequency when the data is sorted according to a selected parameter.

In this way, the databases and methods provided herein permit, among  
5 other things, identification of components, particularly key components, of a disease process by understanding its genetic underpinnings, and also an understanding of processes, such as individual drug responses. The databases and methods provided herein also can be used in methods involving elucidation of pathological pathways, in developing new diagnostic assays, identifying new  
10 potential drug targets, and in identifying new drug candidates.

**Morbidity and/or early mortality associated polymorphisms**

A database containing information provided by a population of healthy blood donors who were not selected for any particular disease to can be used to identify polymorphisms and the alleles in which they are present, whose  
15 frequency decreases with age. These may represent morbidity susceptibility markers and genes.

Polymorphisms of the genome can lead to altered gene function, protein function or genome instability. To identify those polymorphisms which have a clinical relevance/utility is the goal of a world-wide scientific effort. It can be  
20 expected that the discovery of such polymorphisms will have a fundamental impact on the identification and development of novel drug compounds to cure diseases. However, the strategy to identify valuable polymorphisms is cumbersome and dependent upon the availability of many large patient and control cohorts to show disease association. In particular, genes that cause a  
25 general risk of the population to suffer from any disease (morbidity susceptibility genes) will escape these case/control studies entirely.

Here described is a screening strategy to identify morbidity susceptibility genes underlying a variety of different diseases. The definition of a morbidity susceptibility gene is a gene that is expressed in many different cell types or  
30 tissues (housekeeping gene) and its altered function can facilitate the expression of a clinical phenotype caused by disease-specific susceptibility genes that are involved in a pathway specific for this disorder. In other words, these morbidity

-39-

susceptibility genes predispose people to develop a distinct disease according to their genetic make-up for this disease.

Candidates for morbidity susceptibility genes can be found at the bottom level of pathways involving transcription, translation, heat-shock proteins, protein trafficking, DNA repair, assembly systems for subcellular structures (e.g. mitochondria, peroxysomes and other cellular microbodies), receptor signaling cascades, immunology, etc. Those pathways control the quality of life at the cellular level as well as for the entire organism. Mutations/polymorphisms located in genes encoding proteins for those pathways can reduce the fitness of cells and make the organism more susceptible to express the clinical phenotype caused by the action of a disease-specific susceptibility gene. Therefore, these morbidity susceptibility genes can be potentially involved in a whole variety of different complex diseases if not in all. Disease-specific susceptibility genes are involved in pathways that can be considered as disease-specific pathways like glucose-, lipid, hormone metabolism, etc.

The exemplified method permit, among other things, identification of genes and/or gene products involved in a man's general susceptibility to morbidity and/or mortality; use of these genes and/or gene products in studies to elucidate the genetic underpinnings of human diseases; use of these genes and/or gene products in combinatorial statistical analyses without or together with disease-specific susceptibility genes; use of these genes and/or gene products to predict penetrance of disease susceptibility genes; use of these genes and/or gene products in predisposition and/or acute medical diagnostics and use of these genes and/or gene products to develop drugs to cure diseases and/or to extend the life span of humans.

#### SCREENING PROCESS

The healthy population stratified by age, gender and ethnicity, etc. is a very efficient and a universal screening tool for morbidity associated genes. Changes of allelic frequencies in the young compared to the old population are expected to indicate putative morbidity susceptibility genes. Individual samples of this healthy population base can be pooled to further increase the throughput. In a proof of principle experiment pools of young and old Caucasian females and

-40-

males were applied to screen more than 400 randomly chosen single nucleotide polymorphisms located in many different genes. Candidate polymorphisms were identified if the allelic difference was greater than 8% between young and old for both or only one of the genders. The initial results were assayed again in at least one independent subsequent experiments. Repeated experiments are necessary to recognize unstable biochemical reactions, which occur with a frequency of about 2-3% and can mimic age-related allelic frequency differences. Average frequency differences and standard deviations are calculated after successful reproducibility of initial results. The final allelic frequency is then compared to a reference population of Caucasian CEPH sample pool. The result should show similar allelic frequencies in the young Caucasian population. Subsequently, the exact allele frequencies of candidates including genotype information were obtained by analyzing all individual samples. This procedure is straight forward with regard to time and cost. It enables the screening of an enormous number of SNPs. So far, several markers with a highly significant association to age were identified and described below.

In general at least 5 individual in a stratified population need to be screened to produce statistically significant results. The frequency of the allele is determined for an age stratified population. Chi square analysis is then performed on the allelic frequencies to determine if the difference between age groups is statistically significant. A p value less than of 0.1 is considered to represent a statistically significant difference. More preferably the p value should be less than 0.05.

#### Clinical Trials

The identification of markers whose frequency in a population decreases with age also allows for better designed and balanced clinical trials. Currently, if a clinical trial utilizes a marker as a significant endpoint in a study and the marker disappears with age, then the results of the study may be inaccurate. By using methods provided herein, it can be ascertained that if a marker decreases in frequency with age. This information considered and controlled when designing the study. For, example, an age independent marker could be substituted in its place.

-41-

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

#### EXAMPLE 1

This example describes the use of a database containing information provided by a population of healthy blood donors who were not selected for any particular disease to determine the distribution of allelic frequencies of known genetic markers with age and by sex in a Caucasian subpopulation of the database. The results described in this example demonstrate that a disease-related genetic marker or polymorphism can be identified by sorting a healthy database by a parameter or parameters, such as age, sex and ethnicity.

##### Generating a database

Blood was obtained by venous puncture from human subjects who met blood bank criteria for donating blood. The blood samples were preserved with EDTA at pH 8.0 and labeled. Each donor provided information such as age, sex, ethnicity, medical history and family medical history. Each sample was labeled with a barcode representing identifying information. A database was generated by entering, for each donor, the subject identifier and information corresponding to that subject into the memory of a computer storage medium using commercially available software, e.g., Microsoft Access.

##### Model genetic markers

The frequencies of polymorphisms known to be associated at some level with disease were determined in a subpopulation of the subjects represented in the database. These known polymorphisms occur in the p21, p53 and Lipoprotein Lipase genes. Specifically, the N291S polymorphism (N291S) of the Lipoprotein Lipase gene, which results in a substitution of a serine for an asparagine at amino acid codon 291, leads to reduced levels of high density lipoprotein cholesterol (HDL-C) that is associated with an increased risk of males for arteriosclerosis and in particular myocardial infarction (see, Reymer *et al.* (1995) *Nature Genetics* 10:28-34).

The p53 gene encodes a cell cycle control protein that assesses DNA damage and acts as a transcription factor regulating genes that control cell growth, DNA repair and apoptosis (programmed cell death). Mutations in the

-42-

p53 gene have been found in a wide variety of different cancers, including different types of leukemia, with varying frequency. The loss of normal p53 function results in genomic instability and uncontrolled cell growth. A polymorphism that has been identified in the p53 gene, i.e., the R72P allele, results in the substitution of a proline for an arginine at amino acid codon 72 of the gene.

The p21 gene encodes a cyclin-dependent kinase inhibitor associated with G1 phase arrest of normal cells. Expression of the p21 gene triggers apoptosis. Polymorphisms of the p21 gene have been associated with Wilms' tumor, a pediatric kidney cancer. One polymorphism of the p21 gene, the S31R polymorphism, results in a substitution of an arginine for a serine at amino acid codon 31.

#### Database analysis

##### Sorting of subjects according to specific parameters

The genetic polymorphisms were profiled within segments of the Caucasian subpopulation of the sample bank. For p53 profiling, the genomic DNA isolated from blood from a total of 1277 Caucasian subjects age 18-59 years and 457 Caucasian subjects age 60-79 years was analyzed. For p21 profiling, the genomic DNA isolated from blood from a total of 910 Caucasian subjects age 18-49 years and 824 Caucasian subjects age 50-79 years was analyzed. For lipoprotein lipase gene profiling, the genomic DNA from a total of 1464 Caucasian females and 1470 Caucasian males under 60 years of age and a total of 478 Caucasian females and 560 Caucasian males over 60 years of age was analyzed.

##### Isolation and analysis of genomic DNA

Genomic DNA was isolated from blood samples obtained from the individuals. Ten milliliters of whole blood from each individual was centrifuged at 2000 x g. One milliliter of the buffy coat was added to 9 ml of 155 mM  $\text{NH}_4\text{Cl}$ , 10 mM  $\text{KHCO}_3$ , and 0.1 mM  $\text{Na}_2\text{EDTA}$ , incubated 10 min at room temperature and centrifuged for 10 min at 2000 x g. The supernatant was removed, and the white cell pellet was washed in 155 mM  $\text{NH}_4\text{Cl}$ , 10 mM  $\text{KHCO}_3$  and 0.1 mM  $\text{Na}_2\text{EDTA}$  and resuspended in 4.5 ml of 50 mM Tris, 5 mM

-43-

EDTA and 1% SDS. Proteins were precipitated from the cell lysate by 6 mM ammonium acetate, pH 7.3, and then separated from the nucleic acids by centrifugation at 3000 x g. The nucleic acid was recovered from the supernatant by the addition of an equal volume of 100% isopropanol and  
5 centrifugation at 2000 x g. The dried nucleic acid pellet was hydrated in 10 mM Tris, pH 7.6, and 1 mM Na<sub>2</sub>EDTA and stored at 4° C.

Assays of the genomic DNA to determine the presence or absence of the known genetic markers were developed using the BiomassPROBE™ detection method (primer oligo base extension) reaction. This method uses a single  
10 detection primer followed by an oligonucleotide extension step to give products, which can be readily resolved by mass spectrometry, and, in particular, MALDI-TOF mass spectrometry. The products differ in length depending on the presence or absence of a polymorphism. In this method, a detection primer anneals adjacent to the site of a variable nucleotide or sequence of nucleotides  
15 and the primer is extended using a DNA polymerase in the presence of one or more dideoxyNTPs and, optionally, one or more deoxyNTPs. The resulting products are resolved by MALDI-TOF mass spectrometry. The mass of the products as measured by MALDI-TOF mass spectrometry makes possible the determination of the nucleotide(s) present at the variable site.

20 First, each of the Caucasian genomic DNA samples was subjected to nucleic acid amplification using primers corresponding to sites 5' and 3' of the polymorphic sites of the p21 (S31R allele), p53 (R72P allele) and Lipoprotein Lipase (N291S allele) genes. One primer in each primer pair was biotinylated to permit immobilization of the amplification product to a solid support.  
25 Specifically, the polymerase chain reaction primers used for amplification of the relevant segments of the p21, p53 and lipoprotein lipase genes are shown below: US4p21c31-2F (SEQ ID NO: 9) and US5p21-2R (SEQ ID NO: 10) for p21 gene amplification; US4-p53-ex4-F (also shown as p53-ex4US4 (SEQ ID NO: 2)) and US5-p53/2-4R (also shown as US5P53/4R (SEQ ID NO: 3)) for p53 gene  
30 amplification; and US4-LPL-F2 (SEQ ID NO: 16) and US5-LPL-R2 (SEQ ID NO: 17) for lipoprotein lipase gene amplification.

-44-

Amplification of the respective DNA sequences was conducted according to standard protocols. For example, primers may be used in a concentration of 8 pmol. The reaction mixture (e.g., total volume 50  $\mu$ l) may contain

Taq-polymerase including 10x buffer and dTNPs. Cycling conditions for

- 5 polymerase chain reaction amplification may typically be initially 5 min. at 95°C, followed by 1 min. at 94°C, 45 sec at 53°C, and 30 sec at 72°C for 40 cycles with a final extension time of 5 min at 72°C. Amplification products may be purified by using Qiagen's PCR purification kit (No. 28106) according to manufacturer's instructions. The elution of the purified products from the
- 10 column can be done in 50  $\mu$ l TE-buffer (10mM Tris, 1 mM EDTA, pH 7.5).

- The purified amplification products were immobilized via a biotin-avidin linkage to streptavidin-coated beads and the double-stranded DNA was denatured. A detection primer was then annealed to the immobilized DNA using conditions such as, for example, the following: 50  $\mu$ l annealing buffer (20 mM
- 15 Tris, 10 mM KCl, 10 mM  $(\text{NH}_4)_2\text{SO}_4$ , 2 mM  $\text{MgSO}_4$ , 1% Triton X-100, pH 8) at 50°C for 10 min, followed by washing of the beads three times with 200  $\mu$ l washing buffer (40 mM Tris, 1 mM EDTA, 50 mM NaCl, 0.1% Tween 20, pH 8.8) and once in 200  $\mu$ l TE buffer.

- The PROBE extension reaction was performed, for example, by using
- 20 some components of the DNA sequencing kit from USB (No. 70770) and dNTPs or ddNTPs from Pharmacia. An exemplary protocol could include a total reaction volume of 45  $\mu$ l, containing of 21  $\mu$ l water, 6  $\mu$ l Sequenase-buffer, 3  $\mu$ l 10 mM DTT solution, 4.5  $\mu$ l, 0.5 mM of three dNTPs, 4.5  $\mu$ l, 2 mM the missing one ddNTP, 5.5  $\mu$ l glycerol enzyme dilution buffer, 0.25  $\mu$ l Sequenase 2.0, and 0.25
- 25 pyrophosphatase. The reaction can then be pipetted on ice and incubated for 15 min at room temperature and for 5 min at 37°C. The beads may be washed three times with 200  $\mu$ l washing buffer and once with 60  $\mu$ l of a 70 mM  $\text{NH}_4$ -Citrate solution.

- The DNA was denatured to release the extended primers from the
- 30 immobilized template. Each of the resulting extension products was separately analyzed by MALDI-TOF mass spectrometry using 3-hydroxypicolinic acid (3-HPA) as matrix and a UV laser.

-45-

Specifically, the primers used in the PROBE reactions are as shown below: P21/31-3 (SEQ ID NO: 12) for PROBE analysis of the p21 polymorphic site; P53/72 (SEQ ID NO: 4) for PROBE analysis of the p53 polymorphic site; and LPL-2 for PROBE analysis of the lipoprotein lipase gene polymorphic site. In the PROBE analysis of the p21 polymorphic site, the extension reaction was performed using dideoxy-C. The products resulting from the reaction conducted on a "wild-type" allele template (wherein codon 31 encodes a serine) and from the reaction conducted on a polymorphic S31R allele template (wherein codon 31 encodes an arginine) are shown below and designated as P21/31-3 Ser (wt) (SEQ ID NO: 13) and P21/31-3 Arg (SEQ ID NO: 14), respectively. The masses for each product as can be measured by MALDI-TOF mass spectrometry are also provided (i.e., 4900.2 Da for the wild-type product and 5213.4 Da for the polymorphic product).

In the PROBE analysis of the p53 polymorphic site, the extension reaction was performed using dideoxy-C. The products resulting from the reaction conducted on a "wild-type" allele template (wherein codon 72 encodes an arginine) and from the reaction conducted on a polymorphic R72P allele template (wherein codon 72 encodes a proline) are shown below and designated as Cod72 G Arg (wt) and Cod72 C Pro, respectively. The masses for each product as can be measured by MALDI-TOF mass spectrometry are also provided (i.e., 5734.8 Da for the wild-type product and 5405.6 Da for the polymorphic product).

In the PROBE analysis of the lipoprotein lipase gene polymorphic site, the extension reaction was performed using a mixture of ddA and ddT. The products resulting from the reaction conducted on a "wild-type" allele template (wherein codon 291 encodes an asparagine) and from the reaction conducted on a polymorphic N291S allele template (wherein codon 291 encodes a serine) are shown below and designated as 291Asn and 291Ser, respectively. The masses for each product as can be measured by MALDI-TOF mass spectrometry are also provided (i.e., 6438.2 Da for the wild-type product and 6758.4 Da for the polymorphic product).

**P53-1 (R72P)**

**Primers (SEQ ID NOs: 2-4)**

## Masses

**Biotinylated US5 primer is used in the PCR amplification.**

**LPL-1 (N291S)**

25 Amino acid exchange asparagine to serine at codon 291 of the  
lipoprotein lipase gene.

**PCR Product length: 251 bp (SEQ ID NO: 15)**

US4-LPL-F2 (SEQ ID NO: 16)

30 gcgctccatt catctcttca tcgactctct gttgaatgaa gaaaatccaa gtaaggccta  
caggtgcagt tccaaggaag cctttgagaa agggctctgc ttgagttgta gaaagaaccg  
                  LPL-2                   291N  
ctgcaacaat ctgggctatg agatcaataa agtcagagcc aaaagaagca gcaaaatgta  
  g 291S  
35 cctgaagact cgttctcaga tgccc  
                  US4-LPL-R2

**Primers (SEQ ID NOs: 16-18):**

US4-LPL-F2 ccc agt cac gac gtt gta aaa cgg cgc tcc att cat ctc ttc

US5-LPL-R2 agc gga taa caa ttt cac aca ggg ggc atc tga gaa cga gtc

**LPL-2**      **caa tct ggg cta tga qat ca**

-47-

**Masses**

Allele	Product Termination: ddA, ddT	SEQ #	Length	Mass
LPL-2	caatctgggctatgagatca	19	20	6141
291 Asn	caatctgggctatgagatcaa	20	21	6438.2
5 291 Ser	caatctgggctatgagatcagt	21	22	6758.4

Biotinylated US5 primer is used in the PCR amplification.

**P21-1 (S31R)**

Amino acid exchange serine to arginine at codon 31 of the tumor

10 suppressor gene p21. Product length: 207 bp (SEQ ID NO: 8)

US4p21c31-2F

gtcc gtcagaaccc atgcggcagc  
p21/31-3 31S  
15 aaggcctgcc gccgcctctt cggcccagtg gacaqcqagc agctgagccg cgactgtgat  
a 31R  
gcgctaattgg cgggctgcat ccaggaggcc cgtgagcgat ggaacttcga ctttgcacc  
gagacaccac tggaggg  
US5p21-2R

**Primers (SEQ ID NOs: 9-11)**

20 US4p21c31-2F ccc agt cac gac gtt gta aaa cgg tcc gtc aga acc cat gcg g

US5p21-2R agc gga taa caa ttt cac aca ggc tcc agt ggt gtc tcg gtg ac

P21/31-3 cag cga gca gct gag

**Masses**

Allele	Product Termination: ddC	SEQ #	Length	Mass
25 p21/31-3	cagcgagcagctgag	12	15	4627
P21/31-3 Ser (wt)	cagcgagcagctgagc	13	16	4900.2
P21/31-3 Arg	cagcgagcagctgagac	14	17	5213.4

Biotinylated US5 primer is used in the PCR amplification.

30 Each of the Caucasian subject DNA samples was individually analyzed by MALDI-TOF mass spectrometry to determine the identity of the nucleotide at the polymorphic sites. The genotypic results of each assay can be entered into the database. The results were then sorted according to age and/or sex to determine the distribution of allelic

35 frequencies by age and/or sex. As depicted in the Figures showing

-48-

histograms of the results, in each case, there was a differential distribution of the allelic frequencies of the genetic markers for the p21, p53 and lipoprotein lipase gene polymorphisms.

Figure 8 shows the results of the p21 genetic marker assays  
5 reveals a statistically significant decrease (from 13.3% to 9.2%) in the frequency of the heterozygous genotype (S31R) in Caucasians with age (18-49 years of age compared to 50-79 years of age). The frequencies of the homozygous (S31 and R31) genotypes for the two age groups are also shown, as are the overall frequencies of the S31 and R31 alleles in  
10 the two age groups (designated as 'S31 and 'R31, respectively in the Figure).

Figures 7A-C shows the results of the p53 genetic marker assays and reveals a statistically significant decrease (from 6.7% to 3.7%) in the frequency of the homozygous polymorphic genotype (P72) in Caucasians  
15 with age (18-59 years of age compared to 60-79 years of age). The frequencies of the homozygous "wild-type" genotype (R72) and the heterozygous genotype (R72P) for the two age groups are also shown, as are the overall frequencies of the R72 and P72 alleles in the two age groups (designated as 'R72 and 'P72, respectively in the Figure). These  
20 results are consistent with the observation that allele is not benign, as p53 regulates expression of a second protein, p21, which inhibits cyclin-dependent kinases (CDKs) needed to drive cells through the cell-cycle (a mutation in either gene can disrupt the cell cycle leading to increased cell division).

25 Figure 2C shows the results of the lipoprotein lipase gene genetic marker assays reveals a statistically significant decrease (from 1.97% to 0.54%) in the frequency of the polymorphic allele (S291) in Caucasian males with age (see also Reymer *et al.* (1995) *Nature Genetics* 10:28-34).

-49-

The frequencies of this allele in Caucasian females of different age groups are also shown.

## EXAMPLE 2

This example describes the use of MALDI-TOF mass spectrometry  
5 to analyze DNA samples of a number of subjects as individual samples  
and as pooled samples of multiple subjects to assess the presence or  
absence of a polymorphic allele (the 353Q allele) of the Factor VII gene  
and determine the frequency of the allele in the group of subjects. The  
results of this study show that essentially the same allelic frequency can  
10 be obtained by analyzing pooled DNA samples as by analyzing each  
sample separately and thereby demonstrate the quantitative nature of  
MALDI-TOF mass spectrometry in the analysis of nucleic acids.

### Factor VII

Factor VII is a serine protease involved in the extrinsic blood  
15 coagulation cascade. This factor is activated by thrombin and works with  
tissue factor (Factor III) in the processing of Factor X to Factor Xa. There  
is evidence that supports an association between polymorphisms in the  
Factor VII gene and increased Factor VII activity which can result in an  
elevated risk of ischemic cardiovascular disease, including myocardial  
20 infarction. The polymorphism investigated in this study is R353Q (i.e., a  
substitution of a glutamic acid residue for an arginine residue at codon  
353 of the Factor VII gene) (see Table 5).

### Analysis of DNA samples for the presence or absence of the 353Q allele of the Factor VII gene

25 Genomic DNA was isolated from separate blood samples obtained  
from a large number of subjects divided into multiple groups of 92  
subjects per group. Each sample of genomic DNA was analyzed using  
the BiomassPROBE™ assay as described in Example 1 to determine the  
30 presence or absence of the 353Q polymorphism of the Factor VII gene.

-50-

First, DNA from each sample was amplified in a polymerase chain reaction using primers F7-353FUS4 (SEQ ID NO: 24) and F7-353RUS5 (SEQ ID NO: 26) as shown below and using standard conditions, for example, as described in Example 1. One of the primers was biotinylated  
5 to permit immobilization of the amplification product to a solid support. The purified amplification products were immobilized via a biotin-avidin linkage to streptavidin-coated beads and the double-stranded DNA was denatured. A detection primer was then annealed to the immobilized DNA using conditions such as, for example, described in Example 1. The  
10 detection primer is shown as F7-353-P (SEQ ID NO: 27) below. The PROBE extension reaction was carried out using conditions, for example, such as those described in Example 1. The reaction was performed using ddG.

The DNA was denatured to release the extended primers from the  
15 immobilized template. Each of the resulting extension products was separately analyzed by MALDI-TOF mass spectrometry. A matrix such as 3-hydroxypicolinic acid (3-HPA) and a UV laser could be used in the MALDI-TOF mass spectrometric analysis. The products resulting from the reaction conducted on a "wild-type" allele template (wherein codon 353  
20 encodes an arginine) and from the reaction conducted on a polymorphic 353Q allele template (wherein codon 353 encodes a glutamic acid) are shown below and designated as 353 CGG and 353 CAG, respectively. The masses for each product as can be measured by MALDI-TOF mass spectrometry are also provided (i.e., 5646.8 Da for the wild-type product  
25 and 5960 Da for the polymorphic product).

The MALDI-TOF mass spectrometric analyses of the PROBE reactions of each DNA sample were first conducted separately on each sample (250 nanograms total concentration of DNA per analysis). The allelic frequency of the 353Q polymorphism in the group of 92 subjects

-51-

was calculated based on the number of individual subjects in which it was detected.

Next, the samples from 92 subjects were pooled (250 nanograms total concentration of DNA in which the concentration of any individual DNA is 2.7 nanograms) and the pool of DNA was subjected to MALDI-TOF mass spectrometric analysis. The area under the signal corresponding to the mass of the 353Q polymorphism PROBE extension product in the resulting spectrum was integrated in order to quantitate the amount of DNA present. The ratio of this amount to total DNA was used to determine the allelic frequency of the 353Q polymorphism in the group of subjects. This type of individual sample vs. pooled sample analysis was repeated for numerous different groups of 92 different samples.

The frequencies calculated based on individual MALDI-TOF mass spectrometric analysis of the 92 separate samples of each group of 92 are compared to those calculated based on MALDI-TOF mass spectrometric analysis of pools of DNA from 92 samples in Figure 9. These comparisons are shown as "pairs" of bar graphs in the Figure, each pair being labeled as a separate "pool" number, e.g., P1, P16, P2, etc. Thus, for example, for P1, the allelic frequency of the polymorphism calculated by separate analysis of each of the 92 samples was 11.41% and the frequency calculated by analysis of a pool of all of the 92 DNA samples was 12.09%.

The similarity in frequencies calculated by analyzing separate DNA samples individually and by pooling the DNA samples demonstrates that it is possible, through the quantitative nature of MALDI-TOF mass spectrometry, to analyze pooled samples and obtain accurate frequency determinations. The ability to analyze pooled DNA samples significantly reduces the time and costs involved in the use of the non-selected, healthy databases as described herein. It has also been shown that it is

-52-

possible to decrease the DNA concentration of the individual samples in a pooled mixture from 2.7 nanograms to 0.27 nanograms without any change in the quality of the spectrum or the ability to quantitate the amount of sample detected.

### 5 Factor VII R353Q PROBE Assay

PROBE Assay for cod353 CGG>CAG (Arg>Gln), Exon 9 G>A.

PCR fragment: 134 bp (incl. US tags; SEQ ID Nos. 22 and 23)

Frequency of A allele: Europeans about 0.1, Japanese/Chinese about 0.03-0.05 (Thromb. Haemost. 1995, 73:617-22; Diabetologia 1998,

10 41:760-6):

#### F7-353FUS4>

1201 GTGCCGGCTA CTCGGATGGC AGCAAGGACT CCTGCAAGGG GGACAGTGGG  
GGCCCACATG

15 1261 CCACCCACTA CCGGGGCACG TGGTACCTGA CGGGCATCGT CAGCTGGGGC  
CAGGGCTGCG

Primers (SEQ ID Nos: 24-26)

F7-353FUS4 CCC AGT CAC GAC GTT GTA AAA CGA TGG CAG CAA GGA CTC CTG Tm<sup>95</sup> 64°C

F7-353-P CAC ATG CCA CCC ACT ACC

20 F7-353RUS5 AGC GGA TAA CAA TTT CAC ACA GGT GAC GAT GCC CGT CAG GTA C 64°C

#### Masses

Allele	Product Termination: ddG	SEQ #	Length	Mass
F7-353-P	atgccacccactacc	27	18	5333.6
353 CGG	cacatgccacccactaccg	28	19	5646.8
25 353 CAG	cacatgccacccactaccag	29	20	5960
US5-bio bio-	agcggataacaatttcacacagg	30	23	7648.6

#### Conclusion

The above examples demonstrate an effect of altered frequency of  
30 disease causing genetic factors within the general population.

Interpretation of those results allows prediction of the medical relevance of polymorphic genetic alterations. In addition, conclusions can be drawn with regard to their penetrance, diagnostic specificity, positive predictive value, onset of disease, most appropriate onset of preventive strategies,

-53-

and the general applicability of genetic alterations identified in isolated populations to panmixed populations. Therefore, an age- and sex-stratified population-based sample bank that is ethnically homogenous is a suitable tool for rapid identification and validation of genetic factors regarding their potential medical utility.

### EXAMPLE 3

#### MORBIDITY AND MORTALITY MARKERS

##### Sample Band and Initial Screening

Healthy samples were obtained through the blood bank of San Bernardino, CA. Donors signed prior to the blood collection a consent form and agreed that their blood will be used in genetic studies with regard to human aging. All samples were anonymized. Tracking back of samples is not possible.

##### Isolation of DNA from blood samples of a healthy donor population

Blood is obtained from a donor by venous puncture and preserved with 1mM EDTA pH 8.0. Ten milliliters of whole blood from each donor was centrifuged at 2000x g. One milliliter of the buffy coat was added to 9 milliliters of 155mM  $\text{NH}_4\text{Cl}$ , 10mM  $\text{KHCO}_3$ , and 0.1mM  $\text{Na}_2\text{EDTA}$ , incubated 10 minutes at room temperature and centrifuged for 10 minutes at 2000x g. The supernatant was removed, and the white cell pellet was washed in 155mM  $\text{NH}_4\text{Cl}$ , 10mM  $\text{KHCO}_3$ , and 0.1mM  $\text{Na}_2\text{EDTA}$  and resuspended in 4.5 milliliters of 50mM Tris, 5mM EDTA, and 1% SDS. Proteins were precipitated from the cell lysate by 6M Ammonium Acetate, pH 7.3, and separated from the nucleic acid by centrifugation 3000x g. The nucleic acid was recovered from the supernatant by the addition of an equal volume of 100% isopropanol and centrifugation at 2000x g. The dried nucleic acid pellet was hydrated in 10mM Tris pH 7.6 and 1mM  $\text{Na}_2\text{EDTA}$  and stored at 4C.

-54-

In this study, samples were pooled as shown in Table 1. Both parents of the blood donors were of Caucasian origin.

Table 1

Pool ID	Sex	Age-range	# individuals
SP1	Female	18-39 years	276
SP2	Males	18-39 years	276
SP3	Females	60-69 years	184
SP4	Males	60-79 years	368

- 10 More than 400 SNPs were tested using all four pools. After one test run 34 assays were selected to be re-assayed at least once. Finally, 10 assays showed repeatedly differences in allele frequencies of several percent and, therefore, fulfilled the criteria to be tested using the individual samples. Average allele frequency and standard deviation is
- 15 tabulated in Table 2.

Table 2

Assay ID	SP1	SP1-STD	SP2	SP2-STD	SP3	SP3-STD	SP4	SP4-STD
47861	0.457	0.028	0.433	0.042	0.384	0.034	0.380	0.015
47751	0.276	0.007	0.403	0.006	0.428	0.052	0.400	0.097
48319	0.676	0.013	0.627	0.018	0.755	0.009	0.686	0.034
48070	0.581	0.034	0.617	0.045	0.561	n.a.	0.539	0.032
49807	0.504	0.034	0.422	0.020	0.477	0.030	0.556	0.005
49534	0.537	0.017	0.503	n.a.	0.623	0.023	0.535	0.009
49733	0.560	0.006	0.527	0.059	0.546	0.032	0.436	0.016
49947	0.754	0.008	0.763	0.047	0.736	0.052	0.689	0.025
50128	0.401	0.022	0.363	0.001	0.294	0.059	0.345	0.013

63306	0.697	0.012	0.674	0.013	0.712	0.017	0.719	0.005
-------	-------	-------	-------	-------	-------	-------	-------	-------

So far, 7 out of the 10 potential morbidity markers were fully analyzed. Additional information about genes in which these SNPs are  
 5 located was gathered through publicly databases like Genbank.

#### AKAPS

Candidate morbidity and mortality markers include housekeeping genes, such as genes involved in signal transduction. Among such genes are the A-kinase anchoring proteins (AKAPs) genes, which participate in  
 10 signal transduction pathways involving protein phosphorylation. Protein phosphorylation is an important mechanism for enzyme regulation and the transduction of extracellular signals across the cell membrane in eukaryotic cells. A wide variety of cellular substrates, including enzymes, membrane receptors, ion channels and transcription factors, can be  
 15 phosphorylated in response to extracellular signals that interact with cells. A key enzyme in the phosphorylation of cellular proteins in response to hormones and neurotransmitters is cyclic AMP (cAMP)-dependent protein kinase (PKA). Upon activation by cAMP, PKA thus mediates a variety of cellular responses to such extracellular signals. An array of PKA isozymes  
 20 are expressed in mammalian cells. The PKAs usually exist as inactive tetramers containing a regulatory (R) subunit dimer and two catalytic (C) subunits. Genes encoding three C subunits ( $C\alpha$ ,  $C\beta$  and  $C\gamma$ ) and four R subunits ( $R1\alpha$ ,  $R1\beta$ ,  $R11\alpha$  and  $R11\beta$ ) have been identified [see Takio *et al.* (1982) *Proc. Natl. Acad. Sci. U.S. A.* 79:2544-2548; Lee *et al.* (1983) *Proc. Natl. Acad. Sci. U.S. A.* 80:3608-3612; Jahnsen *et al.* (1996) *J. Biol. Chem.* 261:12352-12361; Clegg *et al.* (1988) *Proc. Natl. Acad. Sci. U.S. A.* 85:3703-3707; and Scott (1991) *Pharmacol. Ther.* 50:123-145]. The type I (RI)  $\alpha$  and type II (RII)  $\alpha$  subunits are distributed ubiquitously, whereas  $R1\beta$  and  $R11\beta$  are present mainly in brain [see. *e.g.*, Miki and Eddy

-56-

(1999) *J. Biol. Chem.* 274:29057-29062]. The type I PKA holoenzyme (RI $\alpha$  and RI $\beta$ ) is predominantly cytoplasmic, whereas the majority of type II PKA (RII $\alpha$  and RII $\beta$ ) associates with cellular structures and organelles [Scott (1991) *Pharmacol. Ther.* 50:123-145]. Many hormones and other  
5 signals act through receptors to generate cAMP which binds to the R subunits of PKA and releases and activates the C subunits to phosphorylate proteins. Because protein kinases and their substrates are widely distributed throughout cells, there are mechanisms in place in cells to localize protein kinase-mediated responses to different signals. One  
10 such mechanism involves subcellular targeting of PKAs through association with anchoring proteins, referred to as A-kinase anchoring proteins (AKAPs), that place PKAs in close proximity to specific organelles or cytoskeletal components and particular substrates thereby providing for more specific PKA interactions and localized responses [see,  
15 *e.g.*, Scott *et al.* (1990) *J. Biol. Chem.* 265:21561-21566; Bregman *et al.* (1991) *J. Biol. Chem.* 266:7207-7213; and Miki and Eddy (1999) *J. Biol. Chem.* 274:29057-29062]. Anchoring not only places the kinase close to preferred substrates, but also positions the PKA holoenzyme at sites where it can optimally respond to fluctuations in the second messenger  
20 cAMP [Mochly-Rosen (1995) *Science* 268:247-251; Faux and Scott (1996) *Trends Biochem. Sci.* 21:312-315; Hubbard and Cohen (1993) *Trends Biochem. Sci.* 18:172-177].

Up to 75% of type II PKA is localized to various intracellular sites through association of the regulatory subunit (RII) with AKAPs [see, *e.g.*,  
25 Hausken *et al.* (1996) *J. Biol. Chem.* 271:29016-29022]. RII subunits of PKA bind to AKAPs with nanomolar affinity [Carr *et al.* (1992) *J. Biol. Chem.* 267:13376-13382], and many AKAP-RII complexes have been isolated from cell extracts. RI subunits of PKA bind to AKAPs with only micromolar affinity [Burton *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.*

94:11067-11072]. Evidence of binding of a PKA RI subunit to an AKAP has been reported [Miki and Eddy (1998) *J. Biol. Chem.* 273:34384-34390] in which RI $\alpha$ -specific and RI $\alpha$ /RII $\alpha$  dual specificity PKA anchoring domains were identified on FSC1/AKAP82. Additional dual specific

- 5 AKAPs, referred to as D-AKAP1 and D-AKAP2, which interact with the type I and type II regulatory subunits of PKA have also been reported [Huang *et al.* (1997) *J. Biol. Chem.* 272:8057-8064; Huang *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:11184-11189].

- More than 20 AKAPs have been reported in different tissues and
- 10 species. Complementary DNAs (cDNAs) encoding AKAPs have been isolated from diverse species, ranging from *Caenorhabditis elegans* and *Drosophila* to human [see, *e.g.*, Colledge and Scott (1999) *Trends Cell Biol.* 9:216-221]. Regions within AKAPs that mediate association with RII subunits of PKA have been identified. These regions of approximately
- 15 10-18 amino acid residues vary substantially in primary sequence, but secondary structure predictions indicate that they are likely to form an amphipathic helix with hydrophobic residues aligned along one face of the helix and charged residues along the other [Carr *et al.* (1991) *J. Biol. Chem.* 266:14188-14192; Carr *et al.* (1992) *J. Biol. Chem.* 267:13376-
- 20 13382]. Hydrophobic amino acids with a long aliphatic side chain, *e.g.*, valine, leucine or isoleucine, may participate in binding to RII subunits [Glantz *et al.* (1993) *J. Biol. Chem.* 268:12796-12804].

- Many AKAPs also have the ability to bind to multiple proteins, including other signaling enzymes. For example, AKAP79 binds to PKA,
- 25 protein kinase C (PKC) and the protein phosphatase calcineurin (PP2B) [Coghlan *et al.* (1995) *Science* 267:108-112 and Klauck *et al.* (1996) *Science* 271:1589-1592]. Therefore, the targeting of AKAP79 to neuronal postsynaptic membranes brings together enzymes with opposite catalytic activities in a single complex.

-58-

AKAPs thus serve as potential regulatory mechanisms that increase the selectivity and intensity of a cAMP-mediated response. There is a need, therefore, to identify and elucidate the structural and functional properties of AKAPs in order to gain a complete understanding of the important role these proteins play in the basic functioning of cells.

#### **AKAP10**

The sequence of a human AKAP10 cDNA (also referred to as D-AKAP2) is available in the GenBank database, at accession numbers AF037439 (SEQ ID NO: 31) and NM 007202. The AKAP10 gene is located on chromosome 17.

The sequence of a mouse D-AKAP2 cDNA is also available in the GenBank database (see accession number AF021833). The mouse D-AKAP2 protein contains an RGS domain near the amino terminus that is characteristic of proteins that interact with G $\alpha$  subunits and possess GTPase activating protein-like activity [Huang *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:11184-11189]. The human AKAP10 protein also has sequences homologous to RGS domains. The carboxy-terminal 40 residues of the mouse D-AKAP2 protein are responsible for the interaction with the regulatory subunits of PKA. This sequence is fairly well conserved between the mouse D-AKAP2 and human AKAP10 proteins.

#### **Polymorphisms of the human AKAP10 gene and polymorphic AKAP10 proteins**

Polymorphisms of AKAP genes that alter gene expression, regulation, protein structure and/or protein function are more likely to have a significant effect on the regulation of enzyme (particularly PKA) activity, cellular transduction of signals and responses thereto and on the basic functioning of cells than polymorphisms that do not alter gene and/or protein function. Included in the polymorphic AKAPs provided herein are human AKAP10 proteins containing differing amino acid residues at position number 646.

-59-

Amino acid 646 of the human AKAP10 protein is located in the carboxy-terminal region of the protein within a segment that participates in the binding of R-subunits of PKAs. This segment includes the carboxy-terminal 40 amino acids.

- 5       The amino acid residue reported for position 646 of the human AKAP10 protein is an isoleucine. Polymorphic human AKAP10 proteins provided herein have the amino acid sequence but contain residues other than isoleucine at amino acid position 646 of the protein. In particular  
10       embodiments of the polymorphic human AKAP10 proteins provided herein, the amino acid at position 646 is a valine, leucine or phenylalanine residue.

**An A to G transition at nucleotide 2073 of the human AKAP10 coding sequence**

- 15       As described herein, an allele of the human AKAP10 gene that contains a specific polymorphism at position 2073 of the coding sequence and thereby encodes a valine at position 646 has been detected in varying frequencies in DNA samples from younger and older segments of the human population. In this allele, the A at position 2073 of the AKAP10 gene coding sequence is changed from an A to a G, giving rise  
20       to an altered sequence in which the codon for amino acid 646 changes from ATT, coding for isoleucine, to GTT, coding for valine.

**Morbidity marker 1: human protein kinase A anchoring protein (AKAP10-1)**

- 25       PCR Amplification and BiomassPROBE assay detection of AKAP10-1 in a healthy donor population

PCR Amplification of donor population for AKAP 10

- 30       PCR primers were synthesized by OPERON using phosphoramidite chemistry. Amplification of the AKAP10 target sequence was carried out in single 50 $\mu$ l PCR reaction with 100ng-1ug of pooled human genomic DNAs in a 50 $\mu$ l PCR reaction. Individual DNA concentrations within the

-60-

- pooled samples were present in equal concentration with the final concentration ranging from 1-25ng. Each reaction containing IX PCR buffer (Qiagen, Valencia, CA), 200uM dNTPs, 1U Hotstar Taq polymerase (Qiagen, Valencia, CA), 4mM MgCl<sub>2</sub>, and 25pmol of the
- 5 forward primer containing the universal primer sequence and the target specific sequence 5'-TCTCAATCATGTGCATTGAGG-3'(SEQ ID NO: 45), 2pmol of the reverse primer
- 5'-AGCGGATAACAATTTTCACACAGGGATCACACAGCCATCAGCAG-3' (SEQ ID NO: 46), and 10pmol of a biotinylated universal primer
- 10 complementary to the 5' end of the PCR amplicon
- 5'-AGCGGATAACAATTTTCACACAGG-3'(SEQ ID NO: 47). After an initial round of amplification with the target with the specific forward and reverse primer, the 5' biotinylated universal primer then hybridized and acted as a reverse primer thereby introducing a 3' biotin capture moiety
- 15 into the molecule. The amplification protocol results in a 5'-biotinylated double stranded DNA amplicon and dramatically reduces the cost of high throughput genotyping by eliminating the need to 5' biotin label each forward primer used in a genotyping. Thermal cycling was performed in 0.2mL tubes or 96 well plate using an MJ Research Thermal Cycler
- 20 (calculated temperature) with the following cycling parameters: 94° C for 5 min; 45 cycles: 94° C for 20 sec, 56° C for 30 sec, 72° C for 60 sec; 72° C 3min.

#### Immobilization of DNA

- The 50μl PCR reaction was added to 25ul of streptavidin coated magnetic
- 25 bead (Dyna) prewashed three times and resuspended in 1M NH<sub>4</sub>Cl, 0.06M NH<sub>4</sub>OH. The PCR amplicons were allowed to bind to the beads for 15 minutes at room temperature. The beads were then collected with a magnet and the supernatant containing unbound DNA was removed. The unbound strand was release from the double stranded amplicons by

-61-

incubation in 100mM NaOH and washing of the beads three times with 10mM Tris pH 8.0.

BiomassPROBE assay analysis of donor population for AKAP10-1 (clone 48319)

- 5            Genotyping using the BiomassPROBE assay methods was carried out by resuspending the DNA coated magnetic beads in 26mM Tris-HCl pH 9.5, 6.5 mM MgCl<sub>2</sub> and 50mM each of dTTP and 50mM each of ddCTP, ddATP, ddGTP, 2.5U of a thermostable DNA polymerase (Amersham) and 20pmol of a template specific oligonucleotide PROBE
- 10 primer 5'-CTGGCGCCCCACGTGGTCAA-3' (SEQ ID NO: 48) (Operon). Primer extension occurs with three cycles of oligonucleotide primer hybridization and extension. The extension products were analyzed after denaturation from the template with 50mM NH<sub>4</sub>Cl and transfer of 150nL each sample to a silicon chip preloaded with 150nL of H3PA matrix
- 15 material. The sample material was allowed to crystallize and was analyzed by MALDI-TOF (Bruker, PerSeptive). The SNP that is present in AKAP10-1 is a T to C transversion at nucleotide number 156277 of the sequence of a genomic clone of the AKAP10 gene (GenBank Accession No. AC005730) (SEQ ID NO: 36). SEQ ID NO: 35: represents the
- 20 nucleotide sequence of human chromosome 17, which contains the genomic nucleotide sequence of the human AKAP10 gene, and SEQ ID NO: represents the nucleotide sequence of human chromosome 17, which contains the genomic nucleotide sequence of the human AKAP10-1 allele. The mass of the primer used in the BioMass probe reaction was
- 25 5500.6 daltons. In the presence of the SNP, the primer is extended by the addition of ddC, which has a mass of 5773.8. The wildtype gene results in the addition of dT and ddG to the primer to produce an extension product having a mass of 6101 daltons.

-62-

The frequency of the SNP was measured in a population of age selected healthy individuals. Five hundred fifty-two (552) individuals between the ages of 18-39 years (276 females, 276 males) and 552 individuals between the ages of 60-79 (184 females between the ages of 60-69, 368 males between the age of 60-79) were tested for the presence of the polymorphism localized in the non-translated 3'-region of AKAP 10. Differences in the frequency of this polymorphism with increasing age groups were observed among healthy individuals.

Statistical analysis showed that the significance level for differences in the allelic frequency for alleles between the "younger" and the "older" populations was  $p = 0.0009$  and for genotypes was  $p = 0.003$ .

Differences between age groups are significant. For the total population allele significance is  $p = 0.0009$ , and genotype significance is  $p = 0.003$ .

This marker led to the best significant result with regard to allele and genotype frequencies in the age-stratified population. Figure 19 shows the allele and genotype frequency in both genders as well as in the entire population. For latter the significance for alleles was  $p = 0.0009$  and for genotypes was  $p = 0.003$ . The young and old populations were in Hardy-Weinberg equilibrium. A preferential change of one particular genotype was not seen.

The polymorphism is localized in the non-translated 3'-region of the gene encoding the human protein kinase A anchoring protein (AKAP10). The gene is located on chromosome 17. Its structure includes 15 exons and 14 intervening sequences (introns). The encoded protein is responsible for the sub-cellular localization of the cAMP-dependent protein kinase and, therefore, plays a key role in the G-protein mediated receptor-signaling pathway (Huang et al. PNAS (1007) 94:11184-11189). Since its localization is outside the coding region, this polymorphism is most likely in linkage disequilibrium (LD) with other non-synonymous

-63-

polymorphisms that could cause amino acid substitutions and subsequently alter the function of the protein. Sequence comparison of different Genbank database entries concerning this gene revealed further six potential polymorphisms of which two are supposed to change the  
 5 respective amino acid (see Table 3).

Table 3

Exon	Codon	Nucleotides	Amino acid
3	100	GCT > GCC	Ala > Ala
4	177	AGT > GTG	Met > Val
10 8	424	GGG > GGC	Gly > Gly
10 10	524	CCG > CTG	Pro > Leu
12	591	GTG > GTC	Val > Val
12	599	CGC > CGA	Arg > Arg

#### 15 Morbidity marker 2: human protein kinase A anchoring protein (AKAP10-5)

##### Discovery of AKAP10-5 Allele (SEQ ID NO: 33)

Genomic DNA was isolated from blood (as described above) of seventeen (17) individuals with a genotype CC at the AKAP10-1 gene  
 20 locus and a single heterozygous individual (CT) (as described). A target sequence in the AKAP10-1 gene which encodes the C-terminal PKA binding domain was amplified using the polymerase chain reaction. PCR primers were synthesized by OPERON using phosphoramidite chemistry. Amplification of the AKAP10-1 target sequence was carried out in  
 25 individual 50 $\mu$ l PCR reaction with 25ng of human genomic DNA templates. Each reaction containing 1 X PCR buffer (Qiagen, Valencia, CA), 200 $\mu$ M dNTPs, IU Hotstar Taq polymerase (Qiagen, Valencia, CA), 4mM MgCl<sub>2</sub>, 25pmol of the forward primer (Ex13F) containing the universal primer sequence and the target specific sequence 5'-TCC CAA  
 30 AGT GCT GGA ATT AC-3' (SEQ ID NO: 53), and 2pmol of the reverse

-64-

primer (Ex14R) 5'-GTC CAA TAT ATG CAA ACA GTT G-3' (SEQ ID NO: 54). Thermal cycling was performed in 0.2mL tubes or 96 well plate using an MJ Research Thermal Cycler (MJ Research, Waltham, MA) (calculated temperature) with the following cycling parameters: 94° C for 5 min; 45 cycles; 94° C for 20 sec, 56° C for 30 sec, 72° C for 60 sec; 72° C 3min. After amplification the amplicons were purified using a chromatography (Mo Bio Laboratories (Solana Beach, CA)).

The sequence of the 18 amplicons, representing the target region, was determined using a standard Sanger cycle sequencing method with 25nmol of the PCR amplicon, 3.2uM DNA sequencing primer 5'-CCC ACA GCA GTT AAT CCT TC-3'(SEQ ID NO: 55), and chain terminating dRhodamine labeled 2', 3' dideoxynucleotides (PE Biosystems, Foster City, CA) using the following cycling parameters: 96° C for 15 seconds; 25 cycles: 55° C for 15 seconds, 60° C for 4 minutes. The sequencing products precipitated by 0.3M NaOAc and ethanol. The precipitate was centrifuged and dried. The pellets were resuspended in deionized formamide and separated on a 5% polyacrylimide gel. The sequence was determined using the "Sequencher" software (Gene Codes, Ann Arbor, MI).

The sequence of all 17 of the amplicons, which are homozygous for the AKAP10-1 SNP of the amplicons, revealed a polymorphism at nucleotide position 152171 (numbering for GenBank Accession No. AC005730 for AKAP10 genomic clone (SEQ ID NO: 35)) with A replaced by G. This SNP can also be designated as located at nucleotide 2073 of a cDNA clone of the wildtype AKAP10 (GenBank Accession No. AF037439) (SEQ ID NO: 31). The amino acid sequence of the human AKAP10 protein is provided as SEQ ID NO: 32. This single nucleotide polymorphism was designated as AKAP10-5 (SEQ ID NO: 33) and resulted in a substitution of a valine for an isoleucine residue at amino

-65-

acid position 646 of the amino acid sequence of human AKAP10 (SEQ ID NO: 32).

**PCR Amplification and BiomassPROBE assay detection of AKAP10-5 in a healthy donor population**

5           The healthy population stratified by age is a very efficient and a universal screening tool for morbidity associated genes by allowing for the detection of changes of allelic frequencies in the young compared to the old population. Individual samples of this healthy population base can be pooled to further increase the throughput.

10           Healthy samples were obtained through the blood bank of San Bernardino, CA. Both parents of the blood donors were of Caucasian origin. Practically a healthy subject, when human, is defined as human donor who passes blood bank criteria to donate blood for eventual use in the general population. These criteria are as follows: free of detectable  
15   viral, bacterial, mycoplasma, and parasitic infections; not anemic; and then further selected based upon a questionnaire regarding history (see Figure 3). Thus, a healthy population represents an unbiased population of sufficient health to donate blood according to blood bank criteria, and not further selected for any disease state. Typically such individuals are  
20   not taking any medications.

          PCR primers were synthesized by OPERON using phosphoramidite chemistry. Amplification of the AKAP10 target sequence was carried out in a single 50 $\mu$ l PCR reaction with 100ng- 1 $\mu$ g of pooled human genomic DNAs in a 50 $\mu$ l PCR reaction. Individual DNA concentrations within the  
25   pooled samples were present in equal concentration with the final concentration ranging from 1-25ng. Each reaction contained 1X PCR buffer (Qiagen, Valencia, CA), 200 $\mu$ M dNTPs, 1U Hotstar Taq polymerase (Qiagen, Valencia, CA), 4mM MgCl<sub>2</sub>, and 25pmol of the forward primer containing the universal primer sequence and the target specific  
30   sequence 5'-AGCGGATAACAATTTACACAGGGAGCTAGCTTGGAAGAT

-66-

TGC-3' (SEQ ID NO: 41), 2pmol of the reverse primer 5'-GTCCAATATATGCAAACAGTTG-3' (SEQ ID NO: 54), and 10pmol of a biotinylated universal primer complementary to the 5' end of the PCR amplicon BIO:5'-AGCGGATAACAATTTTCACACAGG-3' (SEQ ID NO: 43).

- 5 After an initial round of amplification with the target with the specific forward and reverse primer, the 5' biotinylated universal primer can then be hybridized and acted as a forward primer thereby introducing a 5' biotin capture moiety into the molecule. The amplification protocol resulted in a 5'-biotinylated double stranded DNA amplicon and
- 10 dramatically reduced the cost of high throughput genotyping by eliminating the need to 5' biotin label every forward primer used in a genotyping.

- Thermal cycling was performed in 0.2mL tubes or 96 well plate using an MJ Research Thermal Cycler (calculated temperature) with the
- 15 following cycling parameters: 94° C for 5 min; 45 cycles: 94° C for 20 sec, 56° C for 30 sec; 72° C for 60 sec; 72° C 3min.

#### **Immobilization of DNA**

- The 50  $\mu$ L PCR reaction was added to 25 $\mu$ L of streptavidin coated magnetic beads (Dyna, Oslo, Norway), which were prewashed three
- 20 times and resuspended in 1M NH<sub>4</sub>Cl, 0.06M NH<sub>4</sub>OH. The 5' end of one strand of the double stranded PCR amplicons were allowed to bind to the beads for 15 minutes at room temperature. The beads were then collected with a magnet and the supernatant containing unbound DNA was removed. The hybridized but unbound strand was released from the
- 25 double stranded amplicons by incubation in 100mM NaOH and washing of the beads three times with 10mM Tris pH 8.0.

#### **Detection of AKAP10-5 using BiomassPROBE™ Assay**

BiomassPROBE™ assay of primer extension analysis (see, U.S. Patent No. 6,043,031) of donor population for AKAP 10-5 (SEQ ID NO:

-67-

33) was performed. Genotyping using these methods was carried out by resuspending the DNA coated magnetic beads in 26mM Tris-HCL pH 9.5, 6.5 mM MgCl<sub>2</sub>, 50mM dTTP, 50mM each of ddCTP, ddATP, ddGTP, 2.5U of a thermostable DNA polymerase (Ambersham), and 20pmol of a

5 template specific oligonucleotide PROBE primer  
 5'-ACTGAGCCTGCTGCATAA-3' (SEQ ID NO: 44) (Operon). Primer extension occurs with three cycles of oligonucleotide primer with hybridization and extension. The extension products were analyzed after denaturation from the template with 50 mM NH<sub>4</sub>Cl and transfer of 150 nL

10 of each sample to a silicon chip preloaded with 150 nl of H3PA matrix material. The sample material was allowed to crystallize and analyzed by MALDI-TOF (Bruker, PerSeptive). The primer has a mass of 5483.6 daltons. The SNP results in the additional of a ddC to the primer, giving a mass of 5756.8 daltons for the extended product. The wild type results in

15 the addition a T and ddG to the primer giving a mass of 6101 daltons.

The frequency of the SNP was measured in a population of age selected healthy individuals. Seven hundred thirteen (713) individuals under 40 years of age (360 females, 353 males) and 703 individuals over 60 years of age (322 females, 381 males) were tested for the presence of

20 the SNP, AKAP10-5 (SEQ ID NO: 33). Results are presented below in Table 1.

TABLE 1					
AKAP10-5 (2073V) frequency comparison in 2 age groups					
			<40	>60	delta G allele
25	Female	Alleles			
		*G	38.6	34.6	4.0
		*A	61.4	65.4	
		Genotypes			
		G	13.9	11.8	2.1
		GA	49.4	45.7	
		A	36.7	42.5	

-68-

5						
	Male	Alleles	*G	41.4	37.0	4.4
			*A	58.6	63.0	
		Genotypes	G	18.4	10.8	7.7
			GA	45.9	52.5	
			A	35.7	36.7	
10						
	Total	Alleles	*G	40.0	35.9	4.1
			*A	60.0	64.1	
		Genotypes	G	16.1	11.2	4.9
			GA	47.7	49.4	
		A	36.2	39.4		

Figure 20 graphically shows these results of allele and genotype distribution in the age and sex stratified Caucasian population.

### Morbidity marker 3: human methionine sulfoxide reductase A (msrA)

The age-related allele and genotype frequency of this marker in both genders and the entire population is shown in Figure 21. The decrease of the homozygous CC genotype in the older male population is highly significant.

### Methionine sulfoxide reductase A (#63306)

PCR Amplification and BiomassPROBE assay detection of the human methionine sulfoxide reductase A (h-msr-A) in a healthy donor population

### PCR Amplification of donor population for h-msr-A

PCR primers were synthesized by OPERON using phosphoramidite chemistry. Amplification of the AKAP10 target sequence was carried out in single 50 $\mu$ l PCR reaction with 100ng-1 $\mu$ g of pooled human genomic DNA templates in a 50 $\mu$ l PCR reaction. Individual DNA concentrations within the pooled samples were present in an equal concentration with

-69-

- the final concentration ranging from 1-25ng. Each reaction containing 1 X PCR buffer (Qiagen, Valencia, CA), 200 $\mu$ M dNTPs, 1U Hotstar Taq polymerase (Qiagen, Valencia, CA), 4mM MgCl<sub>2</sub>, 25pmol of the forward primer containing the universal primer sequence and the target specific
- 5 sequence 5'-TTTCTCTGCACAGAGAGGC-3' (SEQ ID NO: 49), 2pmol of the reverse primer
- 5'-AGCGGATAACAATTTACACAGGGCTGAAATCCTTCGCTTTACC-3' (SEQ ID NO: 50), and 10pmol of a biotinylated universal primer complementary to the 5' end of the PCR amplicon
- 10 5'-AGCGGATAACAATTTACACAGG-3' (SEQ ID NO: 51). After an initial round of amplification of the target with the specific forward and reverse primers, the 5' biotinylated universal primer was then hybridized and acted as a reverse primer thereby introducing a 3' biotin capture moiety into the molecule. The amplification protocol results in a 5'-biotinylated
- 15 double stranded DNA amplicon and and dramatically reduces the cost of high throughput genotyping by eliminating the need to 5' biotin label each forward primer used in a genotyping. Thermal cycling was performed in 0.2mL tubes or 96 well plate using an MJ Research Thermal Cycler (calculated temperature) with the following cycling parameters: 94° C for
- 20 5 min; 45 cycles: 94° C for 20 sec, 56° C for 30 sec, 72° C for 60 sec; 72° C 3min.

#### Immobilization of DNA

- The 50 $\mu$ l PCR reaction was added to 25ul of streptavidin coated magnetic bead (Dynal) prewashed three times and resuspended in 1M
- 25 NH<sub>4</sub>Cl, 0.06M NH<sub>4</sub>OH. The PCR amplicons were allowed to bind to the beads for 15 minutes at room temperature. The beads were then collected with a magnet and the supernatant containing unbound DNA was removed. The unbound strand was release from the double stranded

-70-

amplicons by incubation in 100mM NaOH and washing of the beads three times with 10mM Tris pH 8.0.

**BiomassPROBE assay analysis of donor population for h-msr A**

- Genotyping using the BiomassPROBE assay methods was carried
- 5 out by resuspending the he DNA coated magnetic beads in 26mM Tris-HCl pH 9.5, 6.5 mM MgCl<sub>2</sub>, 50mM of dTTPs and 50mM each of ddCTP, ddATP, ddGTP, 2.5U of a thermostable DNA polymerase (Amersham), and 20pmol of a template specific oligonucleotide PROBE primer 5'-CTGAAAAGGGAGAGAAAG-3' (Operon) (SEQ ID NO: 52).
- 10 Primer extension occurs with three cycles of oligonucleotide primer with hybridization and extension. The extension products were analyzed after denaturation from the template with 50mM NH<sub>4</sub>Cl and transfer of 150nl each sample to a silicon chip preloaded with 150nl of H3PA matrix material. The sample material was allowed to crystallize and analyzed by
- 15 MALDI-TOF (Bruker, PerSeptive). The SNP is represented as a T to C transversion in the sequence of two ESTs. The wild type is represented by having a T at position 128 of GenBank Accession No. AW 195104, which represents the nucleotide sequence of an EST which is a portion of the wild type human msrA gene (SEQ ID NO: 39 ). The SNP is presented
- 20 as a C at position 129 of GenBank Accession No. AW 874187, which represents the nucleotide sequence of an EST which is a portion of an allele of the human msrA gene (SEQ ID NO: 40 ).

In a genomic sequence the SNP is represented as an A to G transversion. The primer utilized in the BioMass probe reaction had a

25 mass of 5654.8 daltons. In the presence of the SNP the primer is extended by the incorporation of a ddC and has a mass of 5928. In the presence of the wildtype the primer is extended by adding a dT and a DDC to produce a mass of 6232.1 daltons.

-71-

The frequency of the SNP was measured in a population of age selected healthy individuals. Five hundred fifty-two (552) individuals between the ages of 18-39 years (276 females, 276 males and 552 individuals between the age of 60-79 (184 females between the ages of 60-69, 368 males between the age of 60-79) were tested for the presence of the polymorphism localized in the nontranslated 3' region of h-msr-A.

Genotype difference between male age group among healthy individuals is significant. For the male population allele significance is  $p = 0.0009$  and genotype significance is  $p = 0.003$ . The age-related allele and genotype frequency of this marker in both genders and the entire population is shown in Figure 21. The decrease of the homozygous CC genotype in the older male population is highly significant.

The polymorphism is localized in the non-translated 3'-region of the gene encoding the human methionine sulfoxide reductase (h-msrA). The exact localization is 451 base pairs downstream the stop codon (TAA). It is very likely that this SNP is in linkage disequilibrium (LD) with another polymorphism more upstream in the coding or promoter region; thus, it is not directly cause morbidity. The enzyme methionine sulfoxide reductase has been proposed to exhibit multiple biological functions. It may serve to repair oxidative protein damage but also play an important role in the regulation of proteins by activation or inactivation of their biological functions (Moskovitz et al. (1990) PNAS 95:14071-14075). It has also been shown that its activity is significantly reduced in brain tissues of Alzheimer patients (Gabbita et al., (1999) J. Neurochem 73:1660-1666). It is scientifically conceivable that proteins involved in the metabolism of reactive oxygen species are associated to disease.

-72-

## CONCLUSION

The use of the healthy population provides for the identification of morbidity markers. The identification of proteins involved in the G-protein coupled signaling transduction pathway or in the detoxification of oxidative stress can be considered as convincing results. Further confirmation and validation of other potential polymorphisms already identified *in silico* in the gene encoding the human protein kinase A anchoring protein could even provide stronger association to morbidity and demonstrate that this gene product is a suitable pharmaceutical or diagnostic target.

## EXAMPLE 4

### MALDI-TOF Mass Spectrometry Analysis

All of the products of the enzyme assays listed below were analyzed by MALDI-TOF mass spectrometry. A diluted matrix solution (0.15  $\mu$ L) containing of 10:1 3-hydroxypicolinic acid:ammonium citrate in 1:1 water:acetonitrile diluted 2.5-fold with water was pipetted onto a SpectroChip (Sequenom, Inc.) and was allowed to crystallize. Then, 0.15  $\mu$ L of sample was added. A linear PerSeptive Voyager DE mass spectrometer or Bruker Biflex MALDI-TOF mass spectrometer, operating in positive ion mode, was used for the measurements. The sample plates were kept at 18.2 kV for 400 nm after each UV laser shot (approximate 250 laser shots total), and then the target voltage was raised to 20 kV. The original spectra were digitized at 500 MHz.

## EXAMPLE 5

### Sample Conditioning

Where indicated in the examples below, the products of the enzymatic digestions were purified with ZipTips (Millipore, Bedford, MA). The ZipTips were pre-wetted with 10  $\mu$ L 50% acetonitrile and equilibrated 4 times with 10  $\mu$ L 0.1 M TEAAc. The oligonucleotide fragments were

-73-

bound to the C18 in the ZipTip material by continuous aspiration and dispensation of each sample into the ZipTip. Each digested oligonucleotide was conditioned by washing with 10  $\mu$ L 0.1 M TEAAc, followed by 4 washing steps with 10  $\mu$ L H<sub>2</sub>O. DNA fragments were eluted from the

- 5 Ziptip with 7  $\mu$ L 50% acetonitrile.

Any method for condition the samples may be employed. Methods for conditioning, which generally is used to increase peak resolution, are well known (see, *e.g.*, International PCT application No. WO 98/20019).

#### EXAMPLE 6

##### 10 DNA Glycosylase-Mediated Sequence Analysis

DNA Glycosylases modifies DNA at each position that a specific nucleobase resides in the DNA, thereby producing abasic sites. In a subsequent reaction with another enzyme, a chemical, or heat, the phosphate backbone at each abasic site can be cleaved.

- 15 The glycosylase utilized in the following procedures was uracil-DNA glycosylase (UDG). Uracil bases were incorporated into DNA fragments in each position that a thymine base would normally occupy by amplifying a DNA target sequence in the presence of uracil. Each uracil substituted DNA amplicon was incubated with UDG, which cleaved each uracil base
- 20 in the amplicon, and was then subjected to conditions that effected backbone cleavage at each abasic site, which produced DNA fragments. DNA fragments were subjected to MALDI-TOF mass spectrometry analysis. Genetic variability in the target DNA was then assessed by analyzing mass spectra.

- 25 Glycosylases specific for nucleotide analogs or modified nucleotides, as described herein, can be substituted for UDG in the following procedures. The glycosylase methods described hereafter, in conjunction with phosphate backbone cleavage and MALDI, can be used to analyze DNA fragments for the purposes of SNP scanning, bacteria

-74-

typing, methylation analysis, microsatellite analysis, genotyping, and nucleotide sequencing and re-sequencing.

#### A. Genotyping

A glycosylase procedure was used to genotype the DNA sequence encoding UCP-2 (Uncoupling Protein 2). The sequence for UCP-2 is deposited in GenBank under accession number AF096289. The sequence variation genotyped in the following procedure was a cytosine (C-allele) to thymine (T-allele) variation at nucleotide position 4790, which results in a alanine to valine mutation at position 55 in the UCP-2 polypeptide.

DNA was amplified using a PCR procedure with a 50  $\mu$ L reaction volume containing of 5 pmol biotinylated primer having the sequence 5'-TGCTTATCCCTGTAGCTACCCTGTCTTGGCCTTGACAGATCCAA-3' (SEQ ID NO: 91), 15 pmol non-biotinylated primer having the sequence 5'-AGCGGATAACAATTTACACAGGCCATCACACCGCGGTACTG-3' (SEQ ID NO: 92), 200  $\mu$ M dATP, 200  $\mu$ M dCTP, 200  $\mu$ M dGTP, 600  $\mu$ M dUTP (to fully replace dTTP), 1.5 mM to 3 mM  $MgCl_2$ , 1 U of HotStarTaq polymerase, and 25 ng of CEPH DNA. Amplification was effected with 45 cycles at an annealing temperature of 56°C.

The amplification product was then immobilized onto a solid support by incubating 50  $\mu$ L of the amplification reaction with 5  $\mu$ L of prewashed Dynabeads for 20 minutes at room temperature. The supernatant was removed, and the beads were incubated with 50  $\mu$ L of 0.1 M NaOH for 5 minutes at room temperature to denature the double-stranded PCR product in such a fashion that single-stranded DNA was linked to the beads. The beads were then neutralized by three washes with 50  $\mu$ L 10 mM TrisHCl (pH 8). The beads were resuspended in 10  $\mu$ L of a 60mM TrisHCl/1mM EDTA (pH 7.9) solution, and 1 U uracil DNA glycosylase was added to the solution for 45 minutes at 37°C to remove uracil nucleotides present in the single-stranded DNA linked to the beads.

-75-

The beads were then washed two times with 25  $\mu$ L of 10 mM TrisHCl (pH 8) and once with 10  $\mu$ L of water. The biotinylated strands were then eluted from the beads with 12  $\mu$ L of 2 M  $\text{NH}_4\text{OH}$  at 60°C for 10 minutes. The backbone of the DNA was cleaved by incubating the samples for 10 min at 95°C (with a closed lid), and ammonia was evaporated from the samples by incubating the samples for 11 min at 80°C.

The cleavage fragments were then analyzed by MALDI-TOF mass spectrometry as described in Example 4. The T-allele generated a unique fragment of 3254 Daltons. The C-allele generated a unique fragment of 4788 Daltons. These fragments were distinguishable in mass spectra. Thus, the above-identified procedure was successfully utilized to genotype individuals heterozygous for the C-allele and T-allele in UCP-2.

#### **B. Glycosylase Analysis Utilizing Pooled DNA Samples**

The glycosylase assay was conducted using pooled samples to detect genetic variability at the UCP-2 locus. DNA of known genotype was pooled from eleven individuals and was diluted to a fixed concentration of 5 ng/ $\mu$ L. The procedure provided in Example 3A was followed using 2 pmol of forward primer having a sequence of 5'-CCCAGTCACGACGTTGTAAAACGTCTTGGCCTTGCAGATCCAAG- 3' (SEQ ID NO: 93) and 15 pmol of reverse primer having the sequence 5'-AGCGGATAACAATTTACACAGGCCATCACACCGCGGTACTG-3' (SEQ ID NO: 94). In addition, 5 pmol of biotinylated primer having the sequence 5'-bioCCCAGTCACGACGTTGTAAAACG 3' (SEQ ID NO: 97) may be introduced to the PCR reaction after about two cycles. The fragments were analyzed via MALDI-TOF mass spectroscopy (Example 4). As determined in Example 3A, the T-allele, which generated a unique fragment of 3254 Daltons, could be distinguished in mass spectra from the C-allele, which generated a unique fragment of 4788 Daltons. Allelic frequency in the pooled samples was quantified by integrating the area

-76-

under each signal corresponding to an allelic fragment. Integration was accomplished by hand calculations using equations well known to those skilled in the art. In the pool of eleven samples, this procedure suggested that 40.9% of the individuals harbored the T allele and 59.09% of the  
5 individuals harbored the C allele.

### C. Glycosylase-Mediated Microsatellite Analysis

A glycosylase procedure was utilized to identify microsatellites of the Bradykinin Receptor 2 (BKR-2) sequence. The sequence for BKR-2 is deposited in GenBank under accession number X86173. BKR-2 includes  
10 a SNP in the promoter region, which is a C to T variation, as well as a SNP in a repeated unit, which is a G to T variation. The procedure provided in Example 3A was utilized to identify the SNP in the promotor region, the SNP in the microsatellite repeat region, and the number of repeated units in the microsatellite region of BKR-2. Specifically, a  
15 forward PCR primer having the sequence 5'-CTCCAGCTGGGCAGGAGTGC-3' (SEQ ID NO: 95) and a reverse primer having the sequence 5'-CACTTCAGTCGCTCCCT-3' (SEQ ID NO: 96) were utilized to amplify BKR-2 DNA in the presence of uracil. The amplicon was fragmented by UDG followed by backbone cleavage. The  
20 cleavage fragments were analyzed by MALDI-TOF mass spectrometry as described in Example 4.

With regard to the SNP in the BKR-2 promotor region having a C to T variation, the C-allele generated a unique fragment having a mass of 7342.4 Daltons and the T-allele generated a unique fragment having a  
25 mass of 7053.2 Daltons. These fragments were distinguishable in mass spectra. Thus, the above-identified procedure was successfully utilized to genotype individuals heterozygous for the C-allele and T-allele in the promotor region of BKR-2.

-77-

With regard to the SNP in the BKR-2 repeat region having a G to T variation, the T-allele generated a unique fragment having a mass of 1784 Daltons, which was readily detected in a mass spectrum. Hence, the presence of the T-allele was indicative of the G to T sequence variation in the repeat region of BKR-2.

In addition, the number of repeat regions was distinguished between individuals having two repeat sequences and individuals having three repeat sequences in BKR-2. The DNA of these individuals did not harbor the G to T sequence variation in the repeat sequence as each repeat sequence contained a G at the SNP locus. The number of repeat regions was determined in individual samples by calculating the area under a signal corresponding to a unique DNA fragment having a mass of 2771.6 Daltons. This signal in spectra generated from individuals having two repeat regions had an area that was thirty-three percent less than the area under the same signal in spectra generated from individuals having three repeat regions. Thus, the procedures discussed above can be utilized to genotype individuals for the number of repeat sequences present in BKR-2.

**D. Bisulfite Treatment Coupled with Glycosylase Digestion**

Bisulfite treatment of genomic DNA can be utilized to analyze positions of methylated cytosine residues within the DNA. Treating nucleic acids with bisulfite deaminates cytosine residues to uracil residues, while methylated cytosine remains unmodified. Thus, by comparing the sequence of a PCR product generated from genomic DNA that is not treated with bisulfite with the sequence of a PCR product generated from genomic DNA that is treated with bisulfite, the degree of methylation in a nucleic acid as well as the positions where cytosine is methylated can be deduced.

-78-

Genomic DNA (2  $\mu$ g) was digested by incubation with 1  $\mu$ L of a restriction enzyme at 37°C for 2 hours. An aliquot of 3 M NaOH was added to yield a final concentration of 0.3M NaOH in the digestion solution. The reaction was incubated at 37°C for 15 minutes followed by  
5 treatment with 5.35M urea, 4.44M bisulfite, and 10mM hydroquinone, where the final concentration of hydroquinone is 0.5 mM.

The sample that was treated with bisulfite (sample A) was compared to the same digestion sample that had not undergone bisulfite treatment (sample B). After sample A was treated with bisulfite as  
10 described above, sample A and sample B were amplified by a standard PCR procedure. The PCR procedure included the step of overlaying each sample with mineral oil and then subjecting the sample to thermocycling (20 cycles of 15 minutes at 55°C followed by 30 seconds at 95°C). The PCR reaction contained four nucleotide bases, C, A, G, and U. The  
15 mineral oil was removed from each sample, and the PCR products were purified with glassmilk. Sodium iodide (3 volumes) and glassmilk (5  $\mu$ L) were added to samples A and B. The samples were then placed on ice for 8 minutes, washed with 420  $\mu$ L cold buffer, centrifuged for 10 seconds, and the supernatant fractions were removed. This process was  
20 repeated twice and then 25  $\mu$ L of water was added. Samples were incubated for 5 minutes at 37 °C, were centrifuged for 20 seconds, and the supernatant fraction was collected, and then this incubation/centrifugation/supernatant fraction collection procedure was repeated. 50  $\mu$ L 0.1 M NaOH was then added to the samples to denature  
25 the DNA. The samples were incubated at room temperature for 5 minutes, washed three times with 50  $\mu$ L of 10 mM TrisHCl (pH 8), and resuspended in 10  $\mu$ L 60mM TrisHCl/1mM EDTA, pH 7.9.

The sequence of PCR products from sample A and sample B were then treated with 2U of UDG (MBI Fermentas) and then subjected to

-79-

- backbone cleavage, as described herein. The resulting fragments from each of sample A and sample B were analyzed by MALDI-TOF mass spectroscopy as described in Example 4. Sample A gave rise to a greater number of fragments than the number of fragments arising from sample
- 5 B, indicative that the nucleic acid harbored at least one methylated cytosine moiety.

### EXAMPLE 7

#### Fen-Ligase-Mediated Haplotyping

- Haplotyping procedures permit the selection of a fragment from one of an
- 10 individual's two homologous chromosomes and to genotype linked SNPs on that fragment. The direct resolution of haplotypes can yield increased information content, improving the diagnosis of any linked disease genes or identifying linkages associated with those diseases. In previous studies, haplotypes were typically reconstructed indirectly through
- 15 pedigree analysis (in cases where pedigrees were available) through laborious and unreliable allele-specific PCR or through single-molecule dilution methods well known in the art.

- A haplotyping procedure was used to determine the presence of two SNPs, referred to as SNP1 and SNP2, located on one strand in a DNA
- 20 sample. The haplotyping procedure used in this assay utilized Fen-1, a site-specific "flap" endonuclease that cleaves DNA "flaps" created by the overlap of two oligonucleotides hybridized to a target DNA strand. The two overlapping oligonucleotides in this example were short arm and long arm allele-specific adaptors. The target DNA was an amplified nucleic
- 25 acid that had been denatured and contained SNP1 and SNP2.

The short arm adaptor included a unique sequence not found in the target DNA. The 3' distal nucleotide of the short arm adaptor was identical to one of the SNP1 alleles. Moreover, the long arm adaptor included two regions: a 3' region complementary to the short arm and a

-80-

5' gene-specific region complementary to the fragment of interest adjacent to the SNP. If there was a match between the adaptor and one of the homologues, the Fen enzyme recognized and cleaved the overlapping flap. The short arm of the adaptor was then ligated to the remainder of the target fragment (minus the SNP site). This ligated fragment was used as the forward primer for a second PCR reaction in which only the ligated homologue was amplified. The second PCR product (PCR2) was then analyzed by mass spectrometry. If there was no match between the adaptors and the target DNA, there was no overlap, no cleavage by Fen-1, and thus no PCR2 product of interest.

If there was more than one SNP in the sequence of interest, the second SNP (SNP2) was found by using an adaptor that was specific for SNP2 and hybridizing the adaptor to the PCR2 product containing the first SNP. The Fen-ligase and amplification procedures were repeated for the PCR2 product containing the first SNP. If the amplified product yielded a second SNP, then SNP1 and SNP2 were on the same fragment.

If the SNP is unknown, then four allele-specific adaptors (e.g. C, G, A, and T) can be used to hybridize with the target DNA. The substrates are then treated with the Fen-ligase protocol, including amplification. The PCR2 products may be analyzed by PROBE, as described herein, to determine which adaptors were hybridized to the DNA target and thus identify the SNPs in the sequence.

A Fen-ligase assay was used to detect two SNPs present in Factor VII. These SNPs are located 814 base pairs apart from each other. SNP1 was located at position 8401 (C to T), and SNP2 was located at 9215 (G to A) (SEQ ID #).

#### A. First Amplification Step

A PCR product (PCR1) was generated for a known heterozygous individual at SNP1, a short distance from the 5' end of the SNP.

-81-

Specifically, a 10  $\mu$ L PCR reaction was performed by mixing 1.5 mM  $MgCl_2$ , 200  $\mu$ M of each dNTP, 0.5 U HotStar polymerase, 0.1  $\mu$ M of a forward primer having the sequence 5'-GCG CTC CTG TCG GTG CCA (SEQ ID NO: 56), 0.1  $\mu$ M of a reverse primer having the sequence 5'-GCC  
5 TGA CTG GTG GGG CCC (SEQ ID NO: 57), and 1 ng of genomic DNA. The annealing temperature was 58°C, and the amplification process yielded fragments that were 861 bp in length.

The PCR1 reaction mixture was divided in half and was treated with an exonuclease 1/SAP mixture (0.22  $\mu$ L mixture/5  $\mu$ L PCR1 reaction)  
10 which contained 1.0  $\mu$ L SAP and 0.1  $\mu$ L exon1. The exonuclease treatment was done for 30 minutes at 37°C and then 20 minutes at 85°C to denature the DNA.

#### B. Adaptor Oligonucleotides

A solution of allele-specific adaptors (C and T), containing of one  
15 long and one short oligonucleotide per adaptor, was prepared. The long arm and short arm oligonucleotides of each adaptor (10  $\mu$ M) were mixed in a 1:1 ratio and heated for 30 seconds at 95°C. The temperature was reduced in 2°C increments to 37°C for annealing. The C-adaptor had a short arm sequence of 5'-CAT GCA TGC ACG GTC (SEQ ID NO: 58) and  
20 a long arm sequence of 5'-CAG AGA GTA CCC CTC GAC CGT GCA TGC ATG (SEQ ID NO: 59). Hence, the long arm of the adaptor was 30 bp (15 bp gene-specific), and the short arm was 15bp. The T-adaptor had a short arm sequence of 5'-CAT GCA TGC ACG GTT (SEQ ID NO: 60) and a long arm sequence of 5'-GTA CGT ACG TGC CAA CTC CCC ATG AGA  
25 GAC (SEQ ID NO: 61). The adaptor could also have a hairpin structure in which the short and long arm are separated by a loop containing of 3 to 10 nucleotides (SEQ ID NO: 118).

#### C. FEN-ligase reaction

-82-

In two tubes (one tube for each allele-specific adaptor per sample) was placed a solution (Solution A) containing of 3.5  $\mu$ l 10 mM 16%PEG/50 mM MOPS, 1.2  $\mu$ l 25 mM  $MgCl_2$ , 1.5  $\mu$ l 10X Ampligase Buffer, and 2.5  $\mu$ l PCR1. Each tube containing Solution A was incubated  
5 at 95°C for 5 minutes to denature the PCR1 product. A second solution (Solution B) containing of 1.65  $\mu$ l Ampligase (Thermostable ligase, Epicentre Technologies), 1.65  $\mu$ l 200ng/ $\mu$ l MFEN (from *Methanococcus jannaschii*), and 3.0  $\mu$ l of an allele specific adaptor (C or T) was prepared. Thus, different variations of Solution B, each variation containing of  
10 different allele-specific adaptors, were made. Solution B was added to Solution A at 95°C and incubated at 55°C for 3 hours. The total reaction volume was 15.0  $\mu$ l per adaptor-specific reaction. For a bi-allelic system, 2 x 15.0  $\mu$ l reactions were required.

The Fen-ligase reaction in each tube was then deactivated by  
15 adding 8.0  $\mu$ l 10 mM EDTA. Then, 1.0  $\mu$ l exoIII/Buffer (70%/30%) solution was added to each sample and incubated 30 minutes at 37°C, 20 minutes at 70°C (to deactivate exoIII), and 5 minutes at 95°C (to denature the sample and dissociate unused adaptor from template). The samples were cooled in an ice slurry and purified on UltraClean PCR  
20 Clean-up (MoBio) spin columns which removed all fragments less than 100 base pairs in length. The fragments were eluted with 50  $\mu$ l  $H_2O$ .

#### D. Second Amplification Step

A second amplification reaction (PCR2) was conducted in each sample tube using the short arm adaptor (C or T) sequence as the forward  
25 primer (minus the SNP1 site). Only the ligated homologue was amplified. A standard PCR reaction was conducted with a total volume of 10.0  $\mu$ l containing of 1X Buffer (final concentration), 1.5 mM final concentration  $MgCl_2$ , 200  $\mu$ M final concentration dNTPs, 0.5 U HotStar polymerase, 0.1  $\mu$ M final concentration forward primer 5'-CAT GCA TGC ACG GT (SEQ ID

-83-

NO: 62), 0.1  $\mu$ M final concentration reverse primer 5'-GCC TGA CTG GTG GGG CCC (SEQ ID NO: 63), and 1.0  $\mu$ l of the purified FEN-ligase reaction solution. The annealing temperature was 58°C. The PCR2 product was analyzed by MALDI TOF mass spectroscopy as described in Example 4.

- 5 The mass spectrum of Fen SNP1 showed a mass of 6084.08 Daltons, representing the C allele.

#### E. Genotyping Additional SNPs

- The second SNP (SNP2) can be found by using an adaptor that is specific for SNP2 and hybridizing that adaptor to the PCR2 product
- 10 containing the first SNP. The Fen-ligase and amplification procedures are repeated for the PCR2 product containing the first SNP. If the amplified product yields a second SNP, then SN1 and SN2 are on the same fragment. The mass spectrum of SNP2, representing the T allele, showed a mass of 6359.88 Daltons.

- 15 This assay can also be performed upon pooled DNA to yield haplotype frequencies as described herein. The Fen-ligase assay can be used to analyze multiplexes as described herein.

### EXAMPLE 8

#### Nickase-Mediated Sequence Analysis

- 20 A DNA nickase, or DNase, was used to recognize and cleave one strand of a DNA duplex. Two nickases used were NY2A nickase and NYS1 nickase (Megabase) which cleave DNA at the following sites:

NY2A: 5'...R AG...3'

3'...Y↓TC...5' where R = A or G and Y = C or T

- 25 NYS1: 5'...↓CC[A/G/T]...3'

3'... GG[T/C/A]...5'.

-84-

**A. Nickase Digestion**

Tris-HCl (10 mM), KCl (10 mM, pH 8.3), magnesium acetate (25 mM), BSA (1 mg/mL), and 6 U of Cvi NY2A or Cvi NYS1 Nickase (Megabase Research) were added to 25 pmol of double-stranded  
5 oligonucleotide template having a sequence of 5'-CGC AGG GTT TCC  
TCG TCG CAC TGG GCA TGT G-3' (SEQ ID NO: 90, Operon, Alameda,  
CA) synthesized using standard phosphoramidite chemistry. With a total  
volume of 20  $\mu$ L, the reaction mixture was incubated at 37°C for 5 hours,  
and the digestion products were purified using ZipTips (Millipore, Bedford,  
10 MA) as described in Example 5. The samples were analyzed by MALTY-  
TOM mass spectroscopy as described in Example 1. The nickase Cvi  
NY2A yielded three fragments with masses 4049.76 Daltons, 5473.14  
Daltons, and 9540.71 Daltons. The Cvi NYS1 nickase yielded fragments  
with masses 2063.18 Daltons, 3056.48 Daltons, 6492.81 Daltons, and  
15 7450.14 Daltons.

**B. Nickase Digestion of Pooled Samples**

DQA (HLA ClassII-DQ Alpha, expected fragment size = 225bp) was  
amplified from the genomic DNA of 100 healthy individuals. DQA was  
amplified using standard PCR chemistry in a reaction having a total  
20 volume of 50  $\mu$ L containing of 10 mM Tris-HCl, 10 mM KCl (pH 8.3), 2.5  
mM MgCl<sub>2</sub>, 200  $\mu$ M of each dNTP, 10 pmol of a forward primer having  
the sequence 5'-GTG CTG CAG GTG TAA ACT TGT ACC AG-3' (SEQ ID  
NO: 64), 10 pmol of a reverse primer having the sequence 5'-CAC GGA  
TCC GGT AGC AGC GGT AGA GTT G-3' (SEQ ID NO: 65), 1 U DNA  
25 polymerase (Stoffel fragment, Perkin Elme r), and 200ng human genomic  
DNA (2ng DNA/individual). The template was denatured at 94°C for 5  
minutes. Thermal cycling was continued with a touch-down program that  
included 45 cycles of 20 seconds at 94°C, 30 seconds at 56°C, 1

-85-

minute at 72°C, and a final extension of 3 minutes at 72°C. The crude PCR product was used in the subsequent nickase reaction.

The unpurified PCR product was subjected to nickase digestion. Tris-HCl (10 mM), KCl (10 mM, pH 8.3), magnesium acetate (25mM),  
5 BSA (1 mg/mL), and 5 U of Cvi NY2A or Cvi NYS1 Nickase (Megabase Research) were added to 25 pmol of the amplified template with a total reaction volume of 20 $\mu$ L. The mixture was then incubated at 37°C for 5 hours. The digestion products were purified with either ZipTips (Millipore, Bedford, MA) as described in Example 5. The samples were analyzed by  
10 MALDI-TOF mass spectroscopy as described in Example 4. This assay can also be used to do multiplexing and standardless genotyping as described herein.

To simplify the nickase mass spectrum, the two complementary strands can be separated after digestion by using a single-stranded  
15 undigested PCR product as a capture probe. This probe (preparation shown below in Example 8C) can be hybridized to the nickase fragments in hybridization buffer containing 200 mM sodium citrate and 1% blocking reagent (Boehringer Mannheim). The reaction is heated to 95°C for 5 minutes and cooled to room temperature over 30 minutes by using a  
20 thermal cycler (PTC-200 DNA engine, MJ Research, Waltham, MA). The capture probe-nickase fragment is immobilized on 140  $\mu$ g of streptavidin-coated magnetic beads. The beads are subsequently washed three times with 70 mM ammonium citrate. The captured single-stranded nickase fragments are eluted by heating to 80°C for 5 minutes in 5  $\mu$ L of 50 mM  
25 ammonium hydroxide.

### C. Preparation of Capture Probe

The capture probe is prepared by amplifying the human  $\beta$ -globin gene (3' end of intron 1 to 5' end of exon 2) via PCR methods in a total volume of 50  $\mu$ L containing of GeneAmp 1XPCR Buffer II, 10 mM Tris-

-86-

HCl, pH 8.3, 50 mM KCl, 2 mM MgCl<sub>2</sub>, 0.2 mM dNTP mix, 10pmol of each primer (forward primer 5'-ACTGGGCATGTGGAGACAG-3'(SEQ ID NO: 66) and biotinylated reverse primer bio5'-GCACTTTCTTGCCATGAG-3'(SEQ ID: 67), 2 U of AmpliTaq Gold, and 200 ng of human genomic

5 DNA. The template is denatured at 94°C for 8 minutes. Thermal cycling is continued with a touch-down program that included 11 cycles of 20 seconds at 94°C, 30 seconds at 64°C, 1 minute at 72°C; and a final extension of 5 minutes at 72°C. The amplicon is purified using UltraClean™ PCR clean-up kit (MO Bio Laboratories, Solano Beach, CA).

10

### EXAMPLE 9

#### Multiplex Type IIS SNP Assay

A Type IIS assay was used to identify human gene sequences with known SNPs. The Type IIS enzyme used in this assay was Fok I which

15 effected double-stranded cleavage of the target DNA. The assay involved the steps of amplification and Fok I treatment of the amplicon. In the amplification step, the primers were designed so that each PCR product of a designated gene target was less than 100 bases such that a Fok I recognition sequence was incorporated at the 5' and 3' end of the

20 amplicon. Therefore, the fragments that were cleaved by Fok I included a center fragment containing the SNP of interest.

Ten human gene targets with known SNPs were analyzed by this assay. Sequences of the ten gene targets, as well as the primers used to amplify the target regions, are found in Table 5. The ten targets were

25 lipoprotein lipase, prothrombin, factor V, cholesterol ester transfer protein (CETP), factor VII, factor XIII, HLA-H exon 2, HLA-H exon 4, methylenetetrahydrofolate reductase (MTHR), and P53 exon 4 codon 72.

Amplification of the ten human gene sequences were carried out in a single 50 µL volume PCR reaction with 20 ng of human genomic DNA

-87-

template in 5 PCR reaction tubes. Each reaction vial contained 1X PCR buffer (Qiagen), 200 $\mu$ M dNTPs, 1U Hotstar Taq polymerase (Qiagen), 4 mM MgCl<sub>2</sub>, and 10pmol of each primer. US8, having sequence of 5'TCAGTCACGACGTT3'(SEQ ID NO: 68), and US9, having sequence of 5'CGGATAACAATTTC3'(SEQ ID NO: 69), were used for the forward and reverse primers respectively. Moreover, the primers were designed such that a Fok I recognition site was incorporated at the 5' and 3' ends of the amplicon. Thermal cycling was performed in 0.2 mL tubes or a 96 well plate using a MJ Research Thermal Cycler (calculated temperature) with the following cycling parameters: 94°C for 5 minutes; 45 cycles: 94°C for 20 seconds, 56°C for 20 seconds, 72°C for 60 seconds; and 72°C for 3 minutes.

Following PCR, the sample was treated with 0.2 U Exonuclease I (Amersham Pharmacia) and S Alkaline Phosphatase (Amersham Pharmacia) to remove the unincorporated primers and dNTPs. Typically, 0.2 U of exonuclease I and SAP were added to 5  $\mu$ L of the PCR sample. The sample was then incubated at 37°C for 15 minutes. Exonuclease I and SAP were then inactivated by heating the sample up to 85°C for 15 minutes. Fok I digestion was performed by adding 2 U of Fok I (New England Biolab) to the 5 uL PCR sample and incubating at 37°C for 30 minutes. Since the Fok I restriction sites are located on both sides of the amplicon, the 5' and 3' cutoff fragments have higher masses than the center fragment containing the SNP. The sample was then purified by anion exchange and analyzed by MALDI-TOF mass spectrometry as described in Example 4. The masses of the gene fragments from this multiplexing experiment are listed in Table 6. These gene fragments were resolved in mass spectra thereby allowing multiplex analysis of sequence variability in these genes.

**Table 5**  
**Genes for Multiplex Type IIS Assay**

30

-88-

5

10

15

Gene	Sequence	Seq. ID No.	Primers	Seq. ID No.
Lipoprotein Lipase (Asn291Ser)	cctttgagaa agggctctgc ttgagttgta gaaagaaccg ctgcaacaat <u>ctgggctatg agatcafa&gt;qltaa agtcagagcc</u> <u>aaaagaagca gcaaaatgta</u>	98-99	5' caatttcacgctggatgcaatct gggctatgagatc 3'	70
			5' caatttcacacagcggaigcttct tttgctctgact 3'	71
Prothrombin	26731 gaattatatt tgtgtttcta aaactatggt <u>tcccaataaa agtgactctc</u> 26781 <u>agc(q&gt;a)agcctc aatgctccca</u> <u>gtgctattca tgggcagctc tctgggctca</u>	100-101	5' tcagtcacgacgttggatgcca <u>taaaagtgactctcagc 3'</u>	72
			5' cggataacaatttcggatgcact <u>gggagcattgaggc 3'</u>	73
Factor V (Arg506Gln)	taataggact acttctaact tgtaagagca <u>gatccctgga caggc(q&gt;a)agga</u>  <u>atacaggat tttgccttg aagtaacctt tcag</u>	102-103	5' tcagtcacgacgttggatgagca <u>gatccctggacaggc 3'</u>	74
			5' cggataacaatttcggatggaca <u>aaatacctgtattcc 3'</u>	75
Cholesterol ester transfer protein (CETP) (I405V)	1261 ctaccatgg gcatttgatt <u>gcagagcage</u> <u>tccgagtcclg&gt;a) tccagagctt</u>  1311 <u>cctgcagtca atgacacgc ctgtgggcat</u> ccctgaggtc atgtctcgta	104-105	5' tcagtcacgacgttggatgcaga <u>gcagctccgagtc 3'</u>	76
			5' <u>cagcggatgatcattggatgcagg</u> <u>aagctctgg 3'</u>	77
Factor VII (R353Q)	1221 agcaaggact cctgcaaggg ggacagtga <u>ggccacatg ccaccacta</u>  1271 <u>ccia&gt;glgggcag tggacctga</u> <u>cgggcatcgt cagctggggc cagggtgcg</u>	106-107	5' tcagtcacgacgttggatgccca <u>catgccaccactac 3'</u>	78
			5' cggataacaatttcggatgcccg <u>tcaggatccacg 3'</u>	79
Factor XIII (V34L)	111 caataactct aatgcagcgg aagatgacct <u>gcccacagtg gagcttcagg</u>  161 <u>gcig&gt;tgtgtgcc ccggggcgtc</u> <u>aacctgcaag gtatgagcat accccccttc</u>	108-109	5' tcagtcacgacgttggatgccca <u>cagtggagcttcag 3'</u>	80
			5' <u>gctcataccttgcaggatgagc</u> 3'	81
HLA-M exon 2 (His63Asp)	361 ttgaagcttt gggctacgtg <u>gatgaccagc</u> <u>tggtcgtgt ctatgat(c&gt;g)at</u>  411 <u>gagagtcgcc gtgtggagcc ccgaactcca</u> <u>tgggtttcca gtagaatttc</u>	110-111	5' tcagtcacgacgttggatgacca <u>gctgttcgtgttc 3'</u>	82
			5' <u>tacatggagttcggggatgcaca</u> <u>cggcagctctc 3'</u>	83
HLA-M exon 4 (Cys282Tyr)	1021 ggataacctt ggctgtaccc cctggggaag <u>agcagagata tacgt(g&gt;a)lccag</u>  1071 <u>gtggagcacc caggcctgga tcagcccttc</u> atttgtatct gggagccctc	112-113	5' tcagtcacgacgttggatgggga <u>agagcagagatatatcgt 3'</u>	84
			5' <u>gaggggctgtatccaggatgggt</u> <u>gctccac 3'</u>	85

-89-

5

Gene	Sequence	Seq. ID No.	Primers	Seq. ID No.
Methylenetetrahydrofolate reductase (MTHFR) (Ala222Val)	761 <u>tgagcactt gaagga gaag gtgtctgcgg</u> gag(c>t) <u>cgattt catcatcacg</u>	114-115	5' tcagtcacgacgttggatgggga agagcagagatatacgt 3'	86
	811 <u>cagcttttct ttgaggctga cacattcttc</u>		5' gaggggctgatccaggatgggt gctccac 3'	87
P53 Exon4 Codon 72 (Arg72Pro)	12101 <u>tccagatgaa gctcccagaa</u> tgccagaggc tgcctccc(g>c)c gtggccctg	116-117	5' gatgaagctcccaggatgccag aggc 3'	88
	12151 <u>caccagcagc tcctacaccg</u> gcggccctg		5' gccgccggtgtaggatgctgctg gtgc 3'	89

-89/a-

**Table 6**  
**The mass of Center Fragments for Ten Different SNP Typing by**  
**IIS Assay**

Gene	LPL( <sup>Asn</sup> 291 <sup>Ser</sup> )		Prothrombin		FV( <sup>Arg</sup> 506 <sup>Gln</sup> )		CETP( <sup>Arg</sup> 405 <sup>Val</sup> )		FVIII( <sup>Arg</sup> 353 <sup>Gly</sup> )		FXIII( <sup>Val</sup> 34)	
Genotype	A	G	G	A	G	A	G	A	G	A	G	T
+ strand mass (Da)	6213	6229	5845	5829	5677	5661	3388	3372	6128	6112	5058	5033
- strand mass (Da)	6129	6114	5949	5964	5472	5487	3437	3452	6174	6189	4916	4940

Gene	H1ah2		H1ah4		MTHR( <sup>Arg</sup> 222 <sup>Val</sup> )		P53exon4( <sup>Arg</sup> 72 <sup>Pro</sup> )	
Genotype	C	G	G	A	C	T	G	C
+ strand mass (Da)	5889	5929	4392	4376	4400	4415	4586	4546
- strand mass (Da)	5836	5796	4319	4334	4368	4352	4724	4764

-90-

**EXAMPLE 10****Exemplary use of parental medical history parameter for stratification of healthy database**

A healthy database can be used to associate a disease state with a specific allele (SNP) that has been found to show a strong association between age and the allele, in particular the homozygous genotype. The method involves using the same healthy database used to identify the age dependent association, however stratification is by information given by the donors about common disorders from which their parents suffered (the donor's familial history of disease). There are three possible answers a donor could give about the health status of their parents: neither were affected, one was affected or both were affected. Only donors above a certain minimum age, depending on the disease, are utilized, as the donors parents must be old enough to have exhibited clinical disease phenotypes. The genotype frequency in each of these groups is determined and compared with each other. If there is an association of the marker in the donor to a disease the frequency of the heterozygous genotype will be increased. The frequency of the homozygous genotype should not increase, as it should be significantly underrepresented in the healthy population.

-91-

**EXAMPLE 11****Method and Device for Identifying a Biological Sample****Description**

In accordance with the present invention, a method and device for  
5 identifying a biological sample is provided. Referring now to FIG. 24, an  
apparatus 10 for identifying a biological sample is disclosed. The apparatus 10  
for identifying a biological sample generally comprises a mass spectrometer 15  
communicating with a computing device 20. In a preferred embodiment, the  
mass spectrometer may be a MALDI-TOF mass spectrometer manufactured by  
10 Bruker-Franzen Analytik GmbH; however, it will be appreciated that other mass  
spectrometers can be substituted. The computing device 20 is preferably a  
general purpose computing device. However, it will be appreciated that the  
computing device could be alternatively configured, for example, it may be  
integrated with the mass spectrometer or could be part of a computer in a larger  
15 network system.

The apparatus 10 for identifying a biological sample may operate as an  
automated identification system having a robot 25 with a robotic arm 27  
configured to deliver a sample plate 29 into a receiving area 31 of the mass  
spectrometer 15. In such a manner, the sample to be identified may be placed  
20 on the plate 29 and automatically received into the mass spectrometer 15. The  
biological sample is then processed in the mass spectrometer to generate data  
indicative of the mass of DNA fragments in the biological sample. This data may  
be sent directly to computing device 20, or may have some preprocessing or  
filtering performed within the mass spectrometer. In a preferred embodiment,  
25 the mass spectrometer 15 transmits unprocessed and unfiltered mass  
spectrometry data to the computing device 20. However, it will be appreciated  
that the analysis in the computing device may be adjusted to accommodate  
preprocessing or filtering performed within the mass spectrometer.

Referring now to FIG. 25, a general method 35 for identifying a biological  
30 sample is shown. In method 35, data is received into a computing device from a  
test instrument in block 40. Preferably the data is received in a raw,  
unprocessed and unfiltered form, but alternatively may have some form of

-92-

filtering or processing applied. The test instrument of a preferred embodiment is a mass spectrometer as described above. However, it will be appreciated that other test instruments could be substituted for the mass spectrometer.

The data generated by the test instrument, and in particular the mass spectrometer, includes information indicative of the identification of the biological sample. More specifically, the data is indicative of the DNA composition of the biological sample. Typically, mass spectrometry data gathered from DNA samples obtained from DNA amplification techniques are noisier than, for example, those from typical protein samples. This is due in part because protein samples are more readily prepared in more abundance, and protein samples are more easily ionizable as compared to DNA samples. Accordingly, conventional mass spectrometer data analysis techniques are generally ineffective for DNA analysis of a biological sample. To improve the analysis capability so that DNA composition data can be more readily discerned, a preferred embodiment uses wavelet technology for analyzing the DNA mass spectrometry data. Wavelets are an analytical tool for signal processing, numerical analysis, and mathematical modeling. Wavelet technology provides a basic expansion function which is applied to a data set. Using wavelet decomposition, the data set can be simultaneously analyzed in the time and frequency domains. Wavelet transformation is the technique of choice in the analysis of data that exhibit complicated time (mass) and frequency domain information, such as MALDI-TOF DNA data. Wavelet transforms as described herein have superior denoising properties as compared to conventional Fourier analysis techniques. Wavelet transformation has proven to be particularly effective in interpreting the inherently noisy MALDI-TOF spectra of DNA samples. In using wavelets, a "small wave" or "scaling function" is used to transform a data set into stages, with each stage representing a frequency component in the data set. Using wavelet transformation, mass spectrometry data can be processed, filtered, and analyzed with sufficient discrimination to be useful for identification of the DNA composition for a biological sample.

Referring again to FIG. 25, the data received in block 40 is denoised in block 45. The denoised data then has a baseline correction applied in block 50.

-93-

A baseline correction is generally necessary as data coming from the test instrument, in particular a mass spectrometer instrument, has data arranged in a generally exponentially decaying manner. This generally exponential decaying arrangement is not due to the composition of the biological sample, but is a result of the physical properties and characteristics of the test instrument, and other chemicals involved in DNA sample preparation. Accordingly, baseline correction substantially corrects the data to remove a component of the data attributable to the test system, and sample preparation characteristics.

After denoising in block 45 and the baseline correction in block 50, a signal remains which is generally indicative of the composition of the biological sample. However, due to the extraordinary discrimination required for analyzing the DNA composition of the biological sample, the composition is not readily apparent from the denoised and corrected signal. For example, although the signal may include peak areas, it is not yet clear whether these "putative" peaks actually represent a DNA composition, or whether the putative peaks are result of a systemic or chemical aberration. Further, any call of the composition of the biological sample would have a probability of error which would be unacceptable for clinical or therapeutic purposes. In such critical situations, there needs to be a high degree of certainty that any call or identification of the sample is accurate. Therefore, additional data processing and interpretation is necessary before the sample can be accurately and confidently identified.

Since the quantity of data resulting from each mass spectrometry test is typically thousands of data points, and an automated system may be set to perform hundreds or even thousands of tests per hour, the quantity of mass spectrometry data generated is enormous. To facilitate efficient transmission and storage of the mass spectrometry data, block 55 shows that the denoised and baseline corrected data is compressed.

In a preferred embodiment, the biological sample is selected and processed to have only a limited range of possible compositions. Accordingly, it is therefore known where peaks indicating composition should be located, if present. Taking advantage of knowing the location of these expected peaks, in block 60 the method 35 matches putative peaks in the processed signal to the

-94-

location of the expected peaks. In such a manner, the probability of each putative peak in the data being an actual peak indicative of the composition of the biological sample can be determined. Once the probability of each peak is determined in block 60, then in block 65 the method 35 statistically determines the composition of the biological sample, and determines if confidence is high enough to calling a genotype.

Referring again to block 40, data is received from the test instrument, which is preferably a mass spectrometer. In a specific illustration, FIG. 26 shows an example of data from a mass spectrometer. The mass spectrometer data 70 generally comprises data points distributed along an x-axis 71 and a y-axis 72. The x-axis 71 represents the mass of particles detected, while the y-axis 72 represents a numerical concentration of the particles. As can be seen in FIG. 26, the mass spectrometry data 70 is generally exponentially decaying with data at the left end of the x-axis 73 generally decaying in an exponential manner toward data at the heavier end 74 of the x-axis 71. However, the general exponential presentation of the data is not indicative of the composition of the biological sample, but is more reflective of systematic error and characteristics. Further, as described above and illustrated in FIG. 26, considerable noise exists in the mass spectrometry DNA data 70.

Referring again to block 45, where the raw data received in block 40 is denoised, the denoising process will be described in more detail. As illustrated in FIG. 25, the denoising process generally entails 1) performing a wavelet transformation on the raw data to decompose the raw data into wavelet stage coefficients; 2) generating a noise profile from the highest stage of wavelet coefficients; and 3) applying a scaled noise profile to other stages in the wavelet transformation. Each step of the denoising process is further described below.

Referring now to FIG. 27, the wavelet transformation of the raw mass spectrometry data is generally diagramed. Using wavelet transformation techniques, the mass spectrometry data 70 is sequentially transformed into stages. In each stage the data is represented in a high stage and a low stage, with the low stage acting as the input to the next sequential stage. For example, the mass spectrometry data 70 is transformed into stage 0 high data

-95-

82 and stage 0 low data 83. The stage 0 low data 83 is then used as an input to the next level transformation to generate stage 1 high data 84 and stage 1 low data 85. In a similar manner, the stage 1 low data 85 is used as an input to be transformed into stage 2 high data 86 and stage 2 low data 87. The transformation is continued until no more useful information can be derived by further wavelet transformation. For example, in the preferred embodiment a 24-point wavelet is used. More particularly a wavelet commonly referred to as the Daubechies 24 is used to decompose the raw data. However, it will be appreciated that other wavelets can be used for the wavelet transformation.

Since each stage in a wavelet transformation has one-half the data points of the previous stage, the wavelet transformation can be continued until the stage n low data 89 has around 50 points. Accordingly, the stage n high 88 would contain about 100 data points. Since the preferred wavelet is 24 points long, little data or information can be derived by continuing the wavelet transformation on a data set of around 50 points.

FIG. 28 shows an example of stage 0 high data 95. Since stage 0 high data 95 is generally indicative of the highest frequencies in the mass spectrometry data, stage 0 high data 95 will closely relate to the quantity of high frequency noise in the mass spectrometry data. In FIG. 29, an exponential fitting formula has been applied to the stage 0 high data 95 to generate a stage 0 noise profile 97. In particular, the exponential fitting formula is in the format  $A_0 + A_1 \text{EXP}(-A_2 m)$ . It will be appreciated that other expedient fitting formulas or other types of curve fits may be used.

Referring now to FIG. 30, noise profiles for the other high stages are determined. Since the later data points in each stage will likely be representative of the level of noise in each stage, only the later data points in each stage are used to generate a standard deviation figure that is representative of the noise content in that particular stage. More particularly, in generating the noise profile for each remaining stage, only the last five percent of the data points in each stage are analyzed to determine a standard deviation number. It will be appreciated that other numbers of points, or alternative methods could be used to generate such a standard deviation figure.

-96-

The standard deviation number for each stage is used with the stage 0 noise profile (the exponential curve) 97 to generate a scaled noise profile for each stage. For example, FIG. 30 shows that stage 1 high data 98 has stage 1 high data 103 with the last five percent of the data points represented by area 99. The points in area 99 are evaluated to determine a standard deviation number indicative of the noise content in stage 1 high data 103. The standard deviation number is then used with the stage 0 noise profile 97 to generate a stage 1 noise profile.

In a similar manner, stage 2 high 100 has stage 2 high data 104 with the last five percent of points represented by area 101. The data points in area 101 are then used to calculate a standard deviation number which is then used to scale the stage 0 noise profile 97 to generate a noise profile for stage 2 data. This same process is continued for each of the stage high data as shown by the stage n high 105. For stage n high 105, stage n high data 108 has the last five percent of data points indicated in area 106. The data points in area 106 are used to determine a standard deviation number for stage n. The stage n standard deviation number is then used with the stage 0 noise profile 97 to generate a noise profile for stage n. Accordingly, each of the high data stages has a noise profile.

FIG. 31 shows how the noise profile is applied to the data in each stage. Generally, the noise profile is used to generate a threshold which is applied to the data in each stage. Since the noise profile is already scaled to adjust for the noise content of each stage, calculating a threshold permits further adjustment to tune the quantity of noise removed. Wavelet coefficients below the threshold are ignored while those above the threshold are retained. Accordingly, the remaining data has a substantial portion of the noise content removed.

Due to the characteristics of wavelet transformation, the lower stages, such as stage 0 and 1, will have more noise content than the later stages such as stage 2 or stage n. Indeed, stage n low data is likely to have little noise at all. Therefore, in a preferred embodiment the noise profiles are applied more aggressively in the lower stages and less aggressively in the later stages. For example, FIG. 31 shows that stage 0 high threshold is determined by multiplying

-97-

the stage 0 noise profile by a factor of four. In such a manner, significant numbers of data points in stage 0 high data 95 will be below the threshold and therefore eliminated. Stage 1 high threshold 112 is set at two times the noise profile for the stage 1 high data, and stage 2 high threshold 114 is set equal to the noise profile for stage 2 high. Following this geometric progression, stage n high threshold 116 is therefore determined by scaling the noise profile for each respective stage n high by a factor equal to  $(1/2^{n-2})$ . It will be appreciated that other factors may be applied to scale the noise profile for each stage. For example, the noise profile may be scaled more or less aggressively to accommodate specific systemic characteristics or sample compositions. As indicated above, stage n low data does not have a noise profile applied as stage n low data 118 is assumed to have little or no noise content. After the scaled noise profiles have been applied to each high data stage, the mass spectrometry data 70 has been denoised and is ready for further processing. A wavelet transformation of the denoised signal results in the sparse data set 120 as shown in FIG. 31.

Referring again to FIG. 25, the mass spectrometry data received in block 40 has been denoised in block 45 and is now passed to block 50 for baseline correction. Before performing baseline correction, the artifacts introduced by the wavelet transformation procedure are preferably removed. Wavelet transformation results vary slightly depending upon which point of the wavelet is used as a starting point. For example, the preferred embodiment uses the 24-point Daubechies-24 wavelet. By starting the transformation at the 0 point of the wavelet, a slightly different result will be obtained than if starting at points 1 or 2 of the wavelet. Therefore, the denoised data is transformed using every available possible starting point, with the results averaged to determine a final denoised and shifted signal. For example, FIG. 33 shows that the wavelet coefficient is applied 24 different times and then the results averaged to generate the final data set. It will be appreciated that other techniques may be used to accommodate the slight error introduced due to wavelet shifting.

The formula 125 is generally indicated in FIG. 33. Once the signal has been denoised and shifted, a denoised and shifted signal 130 is generated as

-98-

shown in FIG. 58. FIG. 34 shows an example of the wavelet coefficient 135 data set from the denoised and shifted signal 130.

FIG. 36 shows that putative peak areas 145, 147, and 149 are located in the denoised and shifted signal 150. The putative peak areas are systematically  
5 identified by taking a moving average along the signal 150 and identifying sections of the signal 150 which exceed a threshold related to the moving average. It will be appreciated that other methods can be used to identify putative peak areas in the signal 150.

Putative peak areas 145, 147 and 149 are removed from the signal 150  
10 to create a peak-free signal 155 as shown in FIG. 37. The peak-free signal 155 is further analyzed to identify remaining minimum values 157, and the remaining minimum values 157 are connected to generate the peak-free signal 155.

FIG. 38 shows a process of using the peak-free signal 155 to generate a baseline 170 as shown in FIG. 39. As shown in block 162, a wavelet  
15 transformation is performed on the peak-free signal 155. All the stages from the wavelet transformation are eliminated in block 164 except for the n low stage. The n low stage will generally indicate the lowest frequency component of the peak-free signal 155 and therefore will generally indicate the system exponential characteristics. Block 166 shows that a signal is reconstructed from the n low  
20 coefficients and the baseline signal 170 is generated in block 168.

FIG. 39 shows a denoised and shifted data signal 172 positioned adjacent a correction baseline 170. The baseline correction 170 is subtracted from the denoised and shifted signal 172 to generate a signal 175 having a baseline correction applied as shown in FIG. 40. Although such a denoised, shifted, and  
25 corrected signal is sufficient for most identification purposes, the putative peaks in signal 175 are not identifiable with sufficient accuracy or confidence to call the DNA composition of a biological sample.

Referring again to FIG. 25, the data from the baseline correction 50 is now compressed in block 55, the compression technique used in a preferred  
30 embodiment is detailed in FIG. 41. In FIG. 41 the data in the baseline corrected data is presented in an array format 182 with x-axis points 183 having an associated data value 184. The x-axis is indexed by the non-zero wavelet

-99-

coefficients, and the associated value is the value of the wavelet coefficient. In the illustrated data example in table 182, the maximum value 184 is indicated to be 1000. Although a particularly advantageous compression technique for mass spectrometry data is shown, it will be appreciated that other compression techniques can be used. Although not preferred, the data may also be stored without compression.

In compressing the data according to a preferred embodiment, an intermediate format 186 is generated. The intermediate format 186 generally comprises a real number having a whole number portion 188 and a decimal portion 190. The whole number portion is the x-axis point 183 while the decimal portion is the value data 184 divided by the maximum data value. For example, in the data 182 a data value "25" is indicated at x-axis point "100". The intermediate value for this data point would be "100.025".

From the intermediate compressed data 186 the final compressed data 195 is generated. The first point of the intermediate data file becomes the starting point for the compressed data. Thereafter each data point in the compressed data 195 is calculated as follows: the whole number portion (left of the decimal) is replaced by the difference between the current and the last whole number. The remainder (right of the decimal) remains intact. For example, the starting point of the compressed data 195 is shown to be the same as the intermediate data point which is "100.025". The comparison between the first intermediate data point "100.025" and the second intermediate data point "150.220" is "50.220". Therefore, "50.220" becomes the second point of the compressed data 195. In a similar manner, the second intermediate point is "150.220" and the third intermediate data point is "500.0001". Therefore, the third compressed data becomes "350.000". The calculation for determining compressed data points is continued until the entire array of data points is converted to a single array of real numbers.

FIG. 42 generally describes the method of compressing mass spectrometry data, showing that the data file in block 201 is presented as an array of coefficients in block 202. The data starting point and maximum is determined as shown in block 203, and the intermediate real numbers are

-100-

calculated in block 204 as described above. With the intermediate data points generated, the compressed data is generated in block 205. The described compression method is highly advantageous and efficient for compressing data sets such as a processed data set from a mass spectrometry instrument. The method is particularly useful for data, such as mass spectrometry data, that uses large numbers and has been processed to have occasional lengthy gaps in x-axis data. Accordingly, an x-y data array for processed mass spectrometry data may be stored with an effective compression rate of 10x or more. Although the compression technique is applied to mass spectrometry data, it will be appreciated that the method may also advantageously be applied to other data sets.

Referring again to FIG. 25, peak heights are now determined in block 60. The first step in determining peak height is illustrated in FIG. 43 where the signal 210 is shifted left or right to correspond with the position of expected peaks. As the set of possible compositions in the biological sample is known before the mass spectrometry data is generated, the possible positioning of expected peaks is already known. These possible peaks are referred to as expected peaks, such as expected peaks 212, 214, and 216. Due to calibration or other errors in the test instrument data, the entire signal may be shifted left or right from its actual position, therefore, putative peaks located in the signal, such as putative peaks 218, 222, and 224 may be compared to the expected peaks 212, 214, and 216, respectively. The entire signal is then shifted such that the putative peaks align more closely with the expected peaks.

Once the putative peaks have been shifted to match expected peaks, the strongest putative peak is identified in FIG. 44. In a preferred embodiment, the strongest peak is calculated as a combination of analyzing the overall peak height and area beneath the peak. For example, a moderately high but wide peak would be stronger than a very high peak that is extremely narrow. With the strongest putative peak identified, such as putative peak 225, a Gaussian 228 curve is fit to the peak 225. Once the Gaussian is fit, the width (W) of the Gaussian is determined and will be used as the peak width for future calculations.

-101-

As generally addressed above, the denoised, shifted, and baseline-corrected signal is not sufficiently processed for confidently calling the DNA composition of the biological sample. For example, although the baseline has generally been removed, there are still residual baseline effects present. These  
5 residual baseline effects are therefore removed to increase the accuracy and confidence in making identifications.

To remove the residual baseline effects, FIG. 45 shows that the putative peaks 218, 222, and 224 are removed from the baseline corrected signal. The peaks are removed by identifying a center line 230, 232, and 234 of the  
10 putative peaks 218, 222, and 224, respectively and removing an area to the left and to the right of the identified center line. For each putative peak, an area equal to twice the width (W) of the Gaussian is removed from the left of the center line, while an area equivalent to 50 daltons is removed from the right of the center line. It has been found that the area representing 50 daltons is  
15 adequate to sufficiently remove the effect of salt adducts which may be associated with an actual peak. Such adducts appear to the right of an actual peak and are a natural effect from the chemistry involved in acquiring a mass spectrum. Although a 50 Dalton buffer has been selected, it will be appreciated that other ranges or methods can be used to reduce or eliminate adduct effects.

20 The peaks are removed and remaining minima 247 located as shown in FIG. 46 with the minima 247 connected to create signal 245. A quartic polynomial is applied to signal 245 to generate a residual baseline 250 as shown in FIG. 47. The residual baseline 250 is subtracted from the signal 225 to generate the final signal 255 as indicated in FIG. 48. Although the residual  
25 baseline is the result of a quartic fit to signal 245, it will be appreciated that other techniques can be used to smooth or fit the residual baseline.

To determine peak height, as shown in FIG. 49, a Gaussian such as Gaussian 266, 268, and 270 is fit to each of the peaks, such as peaks 260, 262, and 264, respectively. Accordingly, the height of the Gaussian is  
30 determined as height 272, 274, and 276. Once the height of each Gaussian peak is determined, then the method of identifying a biological compound 35 can move into the genotyping phase 65 as shown in FIG. 25.

-102-

An indication of the confidence that each putative peak is an actual peak can be discerned by calculating a signal-to-noise ratio for each putative peak. Accordingly, putative peaks with a strong signal-to-noise ratio are generally more likely to be an actual peak than a putative peak with a lower signal-to-noise ratio. As described above and shown in FIG. 50, the height of each peak, such as height 272, 274, and 276, is determined for each peak, with the height being an indicator of signal strength for each peak. The noise profile, such as noise profile 97, is extrapolated into noise profile 280 across the identified peaks. At the center line of each of the peaks, a noise value is determined, such as noise value 282, 283, and 284. With a signal values and a noise values generated, signal-to-noise ratios can be calculated for each peak. For example, the signal-to-noise ratio for the first peak in FIG. 50 would be calculated as signal value 272 divided by noise value 282, and in a similar manner the signal-to-noise ratio of the middle peak in FIG. 50 would be determined as signal 274 divided by noise value 283.

Although the signal-to-noise ratio is generally a useful indicator of the presence of an actual peak, further processing has been found to increase the confidence by which a sample can be identified. For example, the signal-to-noise ratio for each peak in the preferred embodiment is preferably adjusted by the goodness of fit between a Gaussian and each putative peak. It is a characteristic of a mass spectrometer that sample material is detected in a manner that generally complies with a normal distribution. Accordingly, greater confidence will be associated with a putative signal having a Gaussian shape than a signal that has a less normal distribution. The error resulting from having a non-Gaussian shape can be referred to as a "residual error".

Referring to FIG. 51, a residual error is calculated by taking a root mean square calculation between the Gaussian 293 and the putative peak 290 in the data signal. The calculation is performed on data within one width on either side of a center line of the Gaussian. The residual error is calculated as:

where  $G$  is the Gaussian signal value,  $R$  is the putative peak value, and  $N$  is the number of points from  $-W$  to  $+W$ . The calculated residual error is used to generate an adjusted signal-to-noise ratio, as described below.

-103-

An adjusted signal noise ratio is calculated for each putative peak using the formula  $(S/N) * \exp(-1/R)$ , where S/N is the signal-to-noise ratio, and R is the residual error determined above. Although the preferred embodiment calculates an adjusted signal-to-noise ratio using a residual error for each peak, it will be appreciated that other techniques can be used to account for the goodness of fit between the Gaussian and the actual signal.

Referring now to FIG. 52, a probability is determined that a putative peak is an actual peak. In making the determination of peak probability, a probability profile 300 is generated where the adjusted signal-to-noise ratio is the x-axis and the probability is the y-axis. Probability is necessarily in the range between a 0% probability and a 100% probability, which is indicated as 1. Generally, the higher the adjusted signal-to-noise ratio, the greater the confidence that a putative peak is an actual peak.

At some target value for the adjusted signal-to-noise, it has been found that the probability is 100% that the putative peak is an actual peak and can confidently be used to identify the DNA composition of a biological sample. However, the target value of adjusted signal-to-noise ratio where the probability is assumed to be 100% is a variable parameter which is to be set according to application specific criteria. For example, the target signal-to-noise ratio will be adjusted depending upon trial experience, sample characteristics, and the acceptable error tolerance in the overall system. More specifically, for situations requiring a conservative approach where error cannot be tolerated, the target adjusted signal-to-noise ratio can be set to, for example, 10 and higher. Accordingly, 100% probability will not be assigned to a peak unless the adjusted signal-to-noise ratio is 10 or over.

In other situations, a more aggressive approach may be taken as sample data is more pronounced or the risk of error may be reduced. In such a situation, the system may be set to assume a 100% probability with a 5 or greater target signal-to-noise ratio. Of course, an intermediate signal-to-noise ratio target figure can be selected, such as 7, when a moderate risk of error can be assumed. Once the target adjusted signal-to-noise ratio is set for the method,

-104-

then for any adjusted signal-to-noise ratio a probability can be determined that a putative peak is an actual peak.

Due to the chemistry involved in performing an identification test, especially a mass spectrometry test of a sample prepared by DNA amplifications, the allelic ratio between the signal strength of the highest peak and the signal strength of the second (or third and so on) highest peak should fall within an expected ratio. If the allelic ratio falls outside of normal guidelines, the preferred embodiment imposes an allelic ratio penalty to the probability. For example, FIG. 53 shows an allelic penalty 315 which has an x-axis 317 that is the ratio between the signal strength of the second highest peak divided by signal strength of the highest peak. The y-axis 319 assigns a penalty between 0 and 1 depending on the determined allelic ratio. In the preferred embodiment, it is assumed that allelic ratios over 30% are within the expected range and therefore no penalty is applied. Between a ratio of 10% and 30%, the penalty is linearly increased until at allelic ratios below 10% it is assumed the second-highest peak is not real. For allelic ratios between 10% and 30%, the allelic penalty chart 315 is used to determine a penalty 319, which is multiplied by the peak probability determined in FIG. 52 to determine a final peak probability. Although the preferred embodiment incorporates an allelic ratio penalty to account for a possible chemistry error, it will be appreciated that other techniques may be used. Similar treatment will be applied to the other peaks.

With the peak probability of each peak determined, the statistical probability for various composition components may be determined. As an example, in order to determine the probability of each of three possible combinations of two peaks, -- peak G, peak C and combinations GG, CC and GC. FIG. 54 shows an example where a most probable peak 325 is determined to have a final peak probability of 90%. Peak 325 is positioned such that it represents a G component in the biological sample. Accordingly, it can be maintained that there is a 90% probability that G exists in the biological sample. Also in the example shown in FIG. 54, the second highest probability is peak 330 which has a peak probability of 20%. Peak 330 is at a position associated

-105-

with a C composition. Accordingly, it can be maintained that there is a 20% probability that C exists in the biological sample.

With the probability of G existing (90%) and the probability of C existing (20%) as a starting point, the probability of combinations of G and C existing can be calculated. For example, FIG. 54 indicates that the probability of GG existing 329 is calculated as 72%. This is calculated as the probability of GG is equal to the probability of G existing (90%) multiplied by the probability of C not existing (100% - 20%). So if the probability of G existing is 90% and the probability of C not existing is 80%, the probability of GG is 72%.

In a similar manner, the probability of CC existing is equivalent to the probability of C existing (20%) multiplied by the probability of G not existing (100% - 90%). As shown in FIG. 54, the probability of C existing is 20% while the probability of G not existing is 10%, so therefore the probability of CC is only 2%. Finally, the probability of GC existing is equal to the probability of G existing (90%) multiplied by the probability of C existing (20%). So if the probability of G existing is 90% and the probability of C existing is 20%, the probability of GC existing is 18%. In summary form, then, the probability of the composition of the biological sample is:

probability of GG: 72%;

probability of GC: 18%; and

probability of CC: 2%.

Once the probabilities of each of the possible combinations has been determined, FIG. 55 is used to decide whether or not sufficient confidence exists to call the genotype. FIG. 55 shows a call chart 335 which has an x-axis 337 which is the ratio of the highest combination probability to the second highest combination probability. The y-axis 339 simply indicates whether the ratio is sufficiently high to justify calling the genotype. The value of the ratio may be indicated by M 340. The value of M is set depending upon trial data, sample composition, and the ability to accept error. For example, the value M may be set relatively high, such as to a value 4 so that the highest probability must be at least four times greater than the second highest probability before confidence is established to call a genotype. However, if a certain level of error may be

-106-

acceptable, the value of M may be set to a more aggressive value, such as to 3, so that the ratio between the highest and second highest probabilities needs to be only a ratio of 3 or higher. Of course, moderate value may be selected for M when a moderate risk can be accepted. Using the example of FIG. 54, where  
5 the probability of GG was 72% and the probability of GC was 18%, the ratio between 72% and 18% is 4.0, therefore, whether M is set to 3, 3.5, or 4, the system would call the genotype as GG. Although the preferred embodiment uses a ratio between the two highest peak probabilities to determine if a genotype confidently can be called, it will be appreciated that other methods  
10 may be substituted. It will also be appreciated that the above techniques may be used for calculating probabilities and choosing genotypes (or more general DNA patterns) containing of combinations of more than two peaks.

Referring now to FIG. 56, a flow chart is shown generally defining the process of statistically calling genotype described above. In FIG. 56 block 402  
15 shows that the height of each peak is determined and that in block 404 a noise profile is extrapolated for each peak. The signal is determined from the height of each peak in block 406 and the noise for each peak is determined using the noise profile in block 408. In block 410, the signal-to-noise ratio is calculated for each peak. To account for a non-Gaussian peak shape, a residual error is  
20 determined in block 412 and an adjusted signal-to-noise ratio is calculated in block 414. Block 416 shows that a probability profile is developed, with the probability of each peak existing found in block 418. An allelic penalty may be applied in block 420, with the allelic penalty applied to the adjusted peak probability in block 422. The probability of each combination of components is  
25 calculated in block 424 with the ratio between the two highest probabilities being determined in block 426. If the ratio of probabilities exceeds a threshold value then the genotype is called in block 428.

In another embodiment of the invention, the computing device 20 (Fig.  
30 24) supports "standardless" genotyping by identifying data peaks that contain putative SNPs. Standardless genotyping is used, for example, where insufficient information is known about the samples to determine a distribution of expected

-107-

peak locations, against which an allelic penalty as described above can be reliably calculated. This permits the computing device to be used for identification of peaks that contain putative SNPs from data generated by any assay that fragments a targeted DNA molecule. For such standardless

5 genotyping, peaks that are associated with an area under the data curve that deviates significantly from the typical area of other peaks in the data spectrum are identified and their corresponding mass (location along the x-axis) is determined.

More particularly, peaks that deviate significantly from the average area

10 of other peaks in the data are identified, and the expected allelic ratio between data peaks is defined in terms of the ratio of the area under the data peaks. Theoretically, where each genetic loci has the same molar concentration of analyte, the area under each corresponding peak should be the same, thus producing a 1.0 ratio of the peak area between any two peaks. In accordance

15 with the invention, peaks having a smaller ratio relative to the other peaks in the data will not be recognized as peaks. More particularly, peaks having an area ratio smaller than 30% relative to a nominal value for peak area will be assigned an allelic penalty. The mass of the remaining peaks (their location along the x-axis of the data) will be determined based on oligonucleotide standards.

20 Fig. 57 shows a flow diagram representation of the processing by the computing device 20 (Fig. 24) when performing standardless genotyping. In the first operation, represented by the flow diagram box numbered 502, the computing device receives data from the mass spectrometer. Next, the height of each putative peak in the data sample is determined, as indicated by the block

25 504. After the height of each peak in the mass spectrometer data is determined, a de-noise process 505 is performed, beginning with an extrapolation of the noise profile (block 506), followed by finding the noise of each peak (block 508) and calculating the signal to noise ratio for each data sample (block 510). Each of these operations may be performed in accordance

30 with the description above for denoise operations 45 of Fig. 25. Other suitable denoise operations will occur to those skilled in the art.

-108-

The next operation is to find the residual error associated with each data point. This is represented by the block 512 in Figure 57. The next step, block 514, involves calculating an adjusted signal to noise ratio for each identified peak. A probability profile is developed next (block 516), followed by a  
5 determination of the peak probabilities at block 518. In the preferred embodiment, the denoise operations of Fig. 57, comprising block 502 to block 518, comprise the corresponding operations described above in conjunction with Fig. 56 for block 402 through block 418, respectively.

The next action for the standardless genotype processing is to determine  
10 an allelic penalty for each peak, indicated by the block 524. As noted above, the standardless genotype processing of Fig. 57 determines an allelic penalty by comparing area under the peaks. Therefore, rather than compare signal strength ratios to determine an allelic penalty, such as described above for Fig. 53, the standardless processing determines the area under each of the identified peaks  
15 and compares the ratio of those areas. Determining the area under each peak may be computed using conventional numerical analysis techniques for calculating the area under a curve for experimental data.

Thus, the allelic penalty is assigned in accordance with Fig. 58, which shows that no penalty is assigned to peaks having a peak area relative to an  
20 expected average area value that is greater than 0.30 (30%). The allelic penalty is applied to the peak probability value, which may be determined according to the process such as described in Fig. 52. It should be apparent from Fig. 58 that the allelic penalty imposed for peaks below a ratio of 30% is that such peaks will be removed from further measurement and processing. Other penalty  
25 schemes, however, may be imposed in accordance with knowledge about the data being processed, as determined by those skilled in the art.

After the allelic penalty has been determined and applied, the standardless genotype processing compares the location of the remaining putative peaks to oligonucleotide standards to determine corresponding masses  
30 in the processing for block 524. For standardless genotype data, the processing of the block 524 is performed to determine mass and genotype, rather than performing the operations corresponding to block 424, 426, and 428 of Fig. 33.

-109-

Techniques for performing such comparisons and determining mass will be known to those skilled in the art.

In another embodiment, the computing device 20 (Fig. 24) permits the detection and determination of the mass (location along the x-axis of the data) of the sense and antisense strand of fragments generated in the assay. If desired, the computing device may also detect and determine the quantity (area under each peak) of the respective sense and antisense strands, using a similar technique to that described above for standardless genotype processing. The data generated for each type of strand may then be combined to achieve a data redundancy and to thereby increase the confidence level of the determined genotype. This technique obviates primer peaks that are often observed in data from other diagnostic methods, thereby permitting a higher level of multiplexing. In addition, when quantitation is used in pooling experiments, the ratio of the measured peak areas is more reliably calculated than the peak identifying technique, due to data redundancy.

Fig. 23 is a flow diagram that illustrates the processing implemented by the computing device 20 to perform sense and antisense processing. In the first operation, represented by the flow diagram box numbered 602, the computing device receives data from the mass spectrometer. This data will include data for the sense strand and antisense strand of assay fragments. Next, the height of each putative peak in the data sample is determined, as indicated by the block 604. After the height of each peak in the mass spectrometer data is determined, a de-noise process 605 is performed, beginning with an operation that extrapolates the noise profile (block 606), followed by finding the noise of each peak (block 608) and calculating the signal to noise ratio for each data sample (block 610). Each of these operations may be performed in accordance with the description above for the denoise operations 45 of Fig. 25. Other suitable denoise operations will occur to those skilled in the art. The next operation is to find the residual error associated with each data point. This is represented by the block 612 in Figure 36.

After the residual error for the data of the sense strand and antisense strand has been performed, processing to identify the genotypes will be

-110-

performed for the sense strand and also for the antisense strand. Therefore, Fig. 23 shows that processing includes sense strand processing (block 630) and antisense strand processing (block 640). Each block 630, 640 includes processing that corresponds to adjusting the signal to noise ratio, developing a probability profile, determining an allelic penalty, adjusting the peak probability by the allelic penalty, calculating genotype probabilities, and testing genotype probability ratios, such as described above in conjunction with blocks 414 through 426 of Fig. 56. The processing of each block 630, 640 may, if desired, include standardless processing operations such as described above in conjunction with Fig. 57. The standardless processing may be included in place of or in addition to the processing operations of Fig. 56.

After the genotype probability processing is completed, the data from the sense strand and antisense strand processing is combined and compared to expected database values to obtain the benefits of data redundancy as between the sense strand and antisense strand. Those skilled in the art will understand techniques to take advantage of known data redundancies between a sense strand and antisense strand of assay fragments. This processing is represented by the block 650. After the data from the two strands is combined for processing, the genotype processing is performed (block 660) and the genotype is identified.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

-111-

**WHAT IS CLAIMED IS:**

1. A subcollection of samples from a target population, comprising:  
a plurality of samples, wherein the samples are selected from the group  
consisting of blood, tissue, body fluid, cell, seed, microbe, pathogen and  
5 reproductive tissue samples; and  
a symbology on the containers containing the samples, wherein the  
symbology is representative of the source and/or history of each sample,  
wherein:  
the target population is a healthy population that has not been selected  
10 for any disease state;  
the collection comprises samples from the healthy population; and  
the subcollection is obtained by sorting the collection according to  
specified parameters.
2. The subcollection of claim 1, wherein the parameters are selected  
15 from the group consisting of ethnicity, age, gender, height, weight, alcohol  
intake, number of pregnancies, number of live births, vegetarians, type of  
physical activity, state of residence and/or length of residence in a particular  
state, educational level, age of parent at death, cause of parent death, former or  
current smoker, length of time as a smoker, frequency of smoking, occurrence  
20 of a disease in immediate family (parent, siblings, children), use of prescription  
drugs and/or reason therefor, length and/or number of hospital stays and  
exposure to environmental factors.
3. The subcollection of claim 1, wherein the symbology is a bar code.
4. A method of producing a database, comprising:  
25 identifying healthy members of a population;  
obtaining data comprising identifying information and obtaining historical  
information and data relating to the identified members of the population and  
their immediate family;  
entering the data into a database for each member of the population and  
30 associating the member and the data with an indexer.
5. The method of claim 4, further comprising:  
obtaining a body tissue or body fluid sample;

-112-

analyzing the body tissue or body fluid in the sample; and  
entering the results of the analysis for each member into the database  
and associating each result with the indexer representative of each member.

6. A database produced by the method of claim 4.

5       7. A database produced by the method of claim 5.

8. A database, comprising:

datapoints representative of a plurality of healthy organisms from  
whom biological samples are obtained,

10       wherein each datapoint is associated with data representative of  
the organism type and other identifying information.

9.       The database of claim 8, wherein the datapoints are answers to  
questions regarding one or more of a parameters selected from the group  
consisting of ethnicity, age, gender, height, weight, alcohol intake, number of  
pregnancies, number of live births, vegetarians, type of physical activity, state of  
15       residence and/or length of residence in a particular state, educational level, age  
of parent at death, cause of parent death, former or current smoker, length of  
time as a smoker, frequency of smoking, occurrence of a disease in immediate  
family (parent, siblings, children), use of prescription drugs and/or reason  
therefor, length and/or number of hospital stays and exposure to environmental  
20       factors.

10.      The database of claim 9, wherein the organisms are mammals and  
the samples are body fluids or tissues.

11.      The database of claim 9, wherein the samples are selected from  
blood, blood fractions, cells and subcellular organelles.

25       12      The database of claim 8, further comprising,  
phenotypic data from an organism.

13.      The database of claim 12, wherein the data includes one of physical  
characteristics, background data, medical data, and historical data.

30       14.      The database of claim 8, further comprising,  
genotypic data from nucleic acid obtained from an organism.

-113-

15. The database of claim 14, wherein genotypic data includes, genetic markers, non-coding regions, microsatellites, RFLPs, VNTRs, historical data of the organism, medical history, and phenotypic information.

16. The database of claim 8 that is a relational database.

5 17. The database of claim 16, wherein the data are related to an indexer datapoint representative of each organism from whom data is obtained.

18. A method of identifying polymorphisms that are candidate genetic markers, comprising:

identifying a polymorphism; and

10 identifying any pathway or gene linked to the locus of the polymorphism, wherein

the polymorphisms are identified in samples associated with a target population that comprises healthy subjects.

15 19. The method of claim 18, wherein the polymorphism is identified by detecting the presence of target nucleic acids in a sample by a method, comprising the steps of:

a) hybridizing a first oligonucleotide to the target nucleic acid;

b) hybridizing a second oligonucleotide to an adjacent region of the target nucleic acid;

20 c) ligating the hybridized oligonucleotides; and

c) detecting hybridized first oligonucleotide by mass spectrometry as an indication of the presence of the target nucleic acid.

20. The method of claim 18, wherein the polymorphism is identified by detecting target nucleic acids in a sample by a method, comprising the steps of:

25 a) hybridizing a first oligonucleotide to the target nucleic acid and hybridizing a second oligonucleotide to an adjacent region of the target nucleic acid;

b) contacting the hybridized first and second oligonucleotides with a cleavage enzyme to form a cleavage product; and

30 c) detecting the cleavage product by mass spectrometry as an indication of the presence of the target nucleic acid.

-114-

21. The method of claim 20 wherein the samples are from subjects in a healthy database.

22. The method of claim 18, wherein the polymorphism is identified by identifying target nucleic acids in a sample by primer oligo base extension  
5 (probe).

23. The method of 22, wherein primer oligo base extension, comprises:

- a) obtaining a nucleic acid molecule that contains a target nucleotide;
- b) optionally immobilizing the nucleic acid molecule onto a solid support,  
10 to produce an immobilized nucleic acid molecule;
- c) hybridizing the nucleic acid molecule with a primer oligonucleotide that is complementary to the nucleic acid molecule at a site adjacent to the target nucleotide;
- d) contacting the product of step c) with a composition comprising a  
15 dideoxynucleoside triphosphate or a 3'-deoxynucleoside triphosphates and a polymerase, so that only a dideoxynucleoside or 3'-deoxynucleoside triphosphate that is complementary to the target nucleotide is extended onto the primer; and
- e) detecting the extended primer, thereby identifying the target nucleotide.

20 24. The method of claim 23, wherein detection of the extended primer is effected by mass spectrometry, comprising:  
ionizing and volatilizing the product of step d) ; and  
detecting the extended primer by mass spectrometry, thereby identifying the target nucleotide.

25 25. The method of claim 24, wherein;  
samples are presented to the mass spectrometer as arrays on chips; and  
each sample occupies a volume that is about the size of the laser spot projected by the laser in a mass spectrometer used in matrix-assisted laser desorption/ionization (MALDI) spectrometry.

-115-

26. A combination, comprising:

a database containing parameters associated with a datapoint representative of a subject from whom samples are obtained, wherein the subjects are healthy; and

5 an indexed collection of the samples, wherein the index identifies the subject from whom the sample was obtained.

27 The combination of claim 26, wherein the parameter is selected from the group consisting of ethnicity, age, gender, height, weight, alcohol intake, number of pregnancies, number of live births, vegetarians, type of  
10 physical activity, state of residence and/or length of residence in a particular state, educational level, age of parent at death, cause of parent death, former or current smoker, length of time as a smoker, frequency of smoking, occurrence of disease in immediate family (parent, siblings, children), use of prescription drugs and/or reason therefor, length and/or number of hospital stays and  
15 exposure to environmental factors.

28. The combination of claim 26, wherein the database further contains genotypic data for each subject.

29. The combination of claim 26, wherein the samples are blood.

30 A data storage medium, comprising the database of claim 8.

20 31. A computer system, comprising the database of claim 8.

32. A system for high throughput processing of biological samples, comprising:

a process line comprising a plurality of processing stations, each of which performs a procedure on a biological sample contained in a  
25 reaction vessel;

a robotic system that transports the reaction vessel from processing station to processing station;

a data analysis system that receives test results of the process line and automatically processes the test results to make a determination  
30 regarding the biological sample in the reaction vessel;

a control system that determines when the test at each processing station is complete and, in response, moves the reaction vessel to

-116-

the next test station, and continuously processes reaction vessels one after another until the control system receives a stop instruction; and

5 a database of claim 8, wherein the samples tested by the automated process line comprise samples from subjects in the database.

33. The system of claim 32, wherein one of the processing stations comprises a mass spectrometer.

34. The system of claim 32, wherein the data analysis system processes the test results by receiving test data from the mass spectrometer  
10 such that the test data for a biological sample contains one or more signals, whereupon the data analysis system determines the area under the curve of each signal and normalizes the results thereof and obtains a substantially quantitative result representative of the relative amounts of components in the tested sample.

15 35. A method for high throughput processing of biological samples, the method comprising:

transporting a reaction vessel along a system of claim 32, comprising a process line having a plurality of processing stations, each of which performs a procedure on one or more biological samples  
20 contained in the reaction vessel;

determining when the test procedure at each processing station is complete and, in response, moving the reaction vessel to the next processing station;

25 receiving test results of the process line and automatically processing the test results to make a data analysis determination regarding the biological samples in the reaction vessel; and

processing reaction vessels continuously one after another until receiving a stop instruction, wherein the samples tested by the automated process line comprise samples from subjects in the database.

30 36. The method of 35, wherein one of the processing stations comprises a mass spectrometer.

-117-

37. The method of claim 36, wherein the samples are analyzed by a method comprising primer oligo base extension (probe).

38. The method of claim 37, further comprising:

processing the test results by receiving test data from the mass  
5 spectrometer such that the test data for a biological sample contains one or more signals or numerical values representative of signals, whereupon the data analysis system determines the area under the curve of each signal and normalizes the results thereof and obtains a substantially quantitative result representative of the relative amounts of components in the tested sample.

10 39. The method of claim 37, wherein primer oligo base extension, comprises:

a) obtaining a nucleic acid molecule that contains a target nucleotide;

b) optionally immobilizing the nucleic acid molecule onto a solid support, to produce an immobilized nucleic acid molecule;

15 c) hybridizing the nucleic acid molecule with a primer oligonucleotide that is complementary to the nucleic acid molecule at a site adjacent to the target nucleotide;

d) contacting the product of step c) with composition comprising a dideoxynucleoside triphosphate or a 3'-deoxynucleoside triphosphates and a  
20 polymerase, so that only a dideoxynucleoside or 3'-deoxynucleoside triphosphate that is complementary to the target nucleotide is extended onto the primer; and  
e) detecting the primer, thereby identifying the target nucleotide.

40. The method of 39, wherein detection of the extended primer is effected by mass spectrometry, comprising:

25 ionizing and volatilizing the product of step d); and

detecting the extended primer by mass spectrometry, thereby identifying the target nucleotide.

41. The method of claim 36, wherein the target nucleic acids in the sample are detected and/or identified by a method, comprising the steps of:

30 a) hybridizing a first oligonucleotide to the target nucleic acid;

b) hybridizing a second oligonucleotide to an adjacent region of the target nucleic acid;

-118-

c) ligating then hybridized oligonucleotides; and

c) detecting hybridized first oligonucleotide by mass spectrometry as an indication of the presence of the target nucleic acid.

42. The method of claim 36, wherein the target nucleic acids in the sample are detected and/or identified by a method, comprising the steps of:

a) hybridizing a first oligonucleotide to the target nucleic acid and hybridizing a second oligonucleotide to an adjacent region of the target nucleic acid;

b) contacting the hybridized first and second oligonucleotides with a cleavage enzyme to form a cleavage product; and

c) detecting the cleavage product by mass spectrometry as an indication of the presence of the target nucleic acid.

43. A method of producing a database stored in a computer memory, comprising:

identifying healthy members of a population;

obtaining identifying and historical information and data relating to the identified members of the population;

entering the member-related data into the computer memory database for each identified member of the population and associating the member and the data with an indexer.

44. The method of claim 43, further comprising:

obtaining a body tissue or body fluid sample of an identified member;

analyzing the body tissue or body fluid in the sample; and

entering the results of the analysis for each member into the computer memory database and associating each result with the indexer representative of each member.

45. A database produced by the method of claim 43.

46. A database produced by the method of claim 44.

47. The database of claim 8, wherein:

the organisms are selected from among animals, bacteria, fungi, protozoans and parasites and

-119-

each datapoint is associated with parameters representative of the organism type and identifying information.

48. The database of claim 43, further comprising, phenotypic data regarding each subject.

5 49. The database of claim 47 that is a relational database and the parameters are the answers to the questions in the questionnaire.

50. The database of claim 8, further comprising, genotypic data of nucleic acid of the subject, wherein genotypic data includes, but is not limited to, genetic markers, non-coding regions, 10 microsatellites, restriction fragment length polymorphisms (RFLPs), variable number tandem repeats (VNTRs), historical day of the organism, the medical history of the subject, phenotypic information, and other information.

51. A database, comprising data records stored in computer memory, wherein the data records contain information that identifies healthy members of 15 a population, and also contain identifying and historical information and data relating to the identified members.

52. The database of claim 51, further comprising an index value for each identified member that associates each member of the population with the identifying and historical information and data.

20 53. A computer system, comprising the database of claim 51.

54. An automated process line, comprising the database of claim 51.

55. A method for determining a polymorphism that correlates with age, ethnicity or gender, comprising:

identifying a polymorphism; and

25 determining the frequency of the polymorphism with increasing age, with ethnicity or with gender in a healthy population.

56. A method for determining whether a polymorphism correlates with susceptibility to morbidity, early mortality, or morbidity and early mortality, comprising;

30 identifying a polymorphism; and

determining the frequency of the polymorphism with increasing age in a healthy population.

-120-

57. A high throughput method of determining frequencies of genetic variations, comprising:

selecting a healthy target population and a genetic variation to be assessed;

5 pooling a plurality of samples of biopolymers obtained from members of the population,

determining or detecting the biopolymer that comprises the variation by mass spectrometry;

10 obtaining a mass spectrum or a digital representation thereof; and  
determining the frequency of the variation in the population.

58. The method of claim 57, wherein:

the variation is selected from the group consisting of an allelic variation, a post-translational modification, a nucleic modification, a label, a mass modification of a nucleic acid and methylation; and/or

15 the biopolymer is a nucleic acid, a protein, a polysaccharide, a lipid, a small organic metabolite or intermediate, wherein the concentration of biopolymer of interest is the same in each of the samples; and/or

the frequency is determined by assessing the method comprising  
20 determining the area under the peak in the mass spectrum or digital representation thereof corresponding to the mass of the biopolymer comprising the genomic variation.

59. The method of claim 58, wherein the method for determining the frequency is effected by determining the ratio of the signal or the digital representation thereof to the total area of the entire mass spectrum, which is  
25 corrected for background.

60. A method for discovery of a polymorphism in a population, comprising:

sorting the database of claim 8 according to a selected parameter to identify samples that match the selected parameter;

30 isolating a biopolymer from each identified sample;  
optionally pooling each isolated biopolymer;  
optionally amplifying the amount of biopolymer;

-121-

cleaving the pooled biopolymers to produce fragments thereof;

obtaining a mass spectrum of the resulting fragments and comparing the mass spectrum with a control mass spectrum to identify differences between the spectra and thereby identifying any polymorphisms; wherein:

5       the control mass spectrum is obtained from unsorted samples in the collection or samples sorted according to a different parameter.

61. The method of claim 60, wherein cleaving is effected by contacting the biopolymer with an enzyme.

62. The method of claim 61, wherein the enzyme is selected from the  
10   group consisting of nucleotide glycosylase, a nickase and a type IIS restriction enzyme.

63. The method of claim 60, wherein the biopolymer is a nucleic acid or a protein.

64. The method of claim 60, wherein the the mass spectrometric  
15   format is selected from among Matrix-Assisted Laser Desorption/Ionization, Time-of-Flight (MALDI-TOF), Electrospray (ES), IR-MALDI, Ion Cyclotron Resonance (ICR), Fourier Transform and combinations thereof.

65. A method for discovery of a polymorphism in a population, comprising:

20       obtaining samples of body tissue or fluid from a plurality of organisms;

isolating a biopolymer from each sample;

pooling each isolated biopolymer;

optionally amplifying the amount of biopolymer;

cleaving the pooled biopolymers to produce fragments thereof;

25       obtaining a mass spectrum of the resulting fragments;

comparing the frequency of each fragment to identify fragments present in amounts lower than the average frequency, thereby identifying any polymorphisms.

66. The method of claim 65, wherein cleaving is effected by contacting  
30   the biopolymer with an enzyme.

-122-

67. The method of claim 66, wherein the enzyme is selected from the group consisting of nucleotide glycosylase, a nickase and a type IIS restriction enzyme.

68. The method of claim 65, wherein the biopolymer is a nucleic acid  
5 or a protein.

69. The method of claim 65, wherein the the mass spectrometric format is selected from among Matrix-Assisted Laser Desorption/Ionization, Time-of-Flight (MALDI-TOF), Electrospray (ES), IR-MALDI, Ion Cyclotron Resonance (ICR), Fourier Transform and combinations thereof.

10 70. The method of claim 65, wherein the samples are obtained from healthy subjects.

71. A method of correlating a polymorphism with a parameter, comprising:

sorting the database of claim 8 according to a selected parameter to  
15 identify samples that match the selected parameter;  
isolating a biopolymer from each identified sample;  
pooling each isolated biopolymer;  
optionally amplifying the amount of biopolymer;  
determining the frequency of the polymorphism in the pooled  
20 biopolymers, wherein:  
an alteration of the frequency of the polymorphism compared to a control, indicates a correlation of the polymorphism with the selected parameter; and  
the control is the frequency of the polymorphism in pooled biopolymers obtained from samples identified from an unsorted database or from a database  
25 sorting according to a different parameter.

72. The method claim 71, wherein the parameter is selected from the group consisting of ethnicity, age, gender, height, weight, alcohol intake, number of pregnancies, number of live births, vegetarians, type of physical activity, state of residence and/or length of residence in a particular state,  
30 educational level, age of parent at death, cause of parent death, former or current smoker, length of time as a smoker, frequency of smoking, occurrence of a disease in immediate family (parent, siblings, children), use of prescription

-123-

drugs and/or reason therefor, length and/or number of hospital stays and exposure to environmental factors.

73. The method claim 72, wherein the parameter is occurrence of disease or a particular disease in an immediate family member, thereby  
5 correlating the polymorphism with the disease.

74. The method of claim 71, wherein the pooled biopolymers are pooled nucleic acid molecules.

75. The method of claim 74, wherein the polymorphism is detected by primer oligo base extension (PROBE).

10 76. The method of 75, wherein primer oligo base extension, comprises:

a) optionally immobilizing the nucleic acid molecules onto a solid support, to produce immobilized nucleic acid molecules;

b) hybridizing the nucleic acid molecules with a primer oligonucleotide  
15 that is complementary to the nucleic acid molecule at a site adjacent to the polymorphism;

c) contacting the product of step c) with composition comprising a dideoxynucleoside triphosphate or a 3'-deoxynucleoside triphosphates and a polymerase, so that only a dideoxynucleoside or 3'-deoxynucleoside triphosphate  
20 that is complementary to the polymorphism is extended onto the primer; and

d) detecting the extended primer, thereby detecting the polymorphism in nucleic acid molecules in the pooled nucleic acids.

77. The method of claim 76, wherein detecting is effected by mass spectrometry.

25 78. The method of claim 71, wherein the frequency is percentage of nucleic acid molecules in the pooled nucleic acids that contain the polymorphism.

79. The method of claim 78, wherein the ratio is determined by obtaining mass spectra of the pooled nucleic acids.

30 80. The method of claim 72, wherein the parameter is age, thereby correlating the polymorphism with suceptibility to morbidity, early mortality or morbidity and early mortality.

-124-

81. A method for haplotyping polymorphisms in a nucleic acid, comprising:

- (a) sorting the database of claim 8 according to a selected parameter to identify samples that match the selected parameter;
- 5 (b) isolating nucleic acid from each identified sample;
- (c) optionally pooling each isolated nucleic acid;
- (d) amplifying the amount of nucleic acid;
- (e) forming single-stranded nucleic acid and splitting each single-strand into a separate reaction vessel;
- 10 (f) contacting each single-stranded nucleic acid with an adaptor nucleic acid to form an adaptor complex;
- (g) contacting the adaptor complex with a nuclease and a ligase;
- (h) contacting the products of step (g) with a mixture that is capable of amplifying a ligated adaptor to produce an extended product;
- 15 (i) obtaining a mass spectrum of each nucleic acid resulting from step (h) and detecting a polymorphism by identifying a signal corresponding to the extended product;
- (j) repeating steps (f) through (i) utilizing an adaptor nucleic acid able to hybridize with another adapter nucleic acid that hybridizes to a different
- 20 sequence on the same strand; whereby
- the polymorphisms are haplotyped by detecting more than one extended product.

82. The method of claim 1, wherein the nuclease is Fen-1.

83. A method for haplotyping polymorphisms in a population, comprising:

- sorting the database of claim 8 according to a selected parameter to identify samples that match the selected parameter;
- isolating a nucleic acid from each identified sample;
- pooling each isolated nucleic acid;
- 30 optionally amplifying the amount of nucleic acid;
- contacting the nucleic acid with at least one enzyme to produce fragments thereof;

-125-

obtaining a mass spectrum of the resulting fragments; whereby:

the polymorphisms are detected by detecting signals corresponding to the polymorphisms; and

the polymorphisms are haplotyped by determining from the mass  
5 spectrum that the polymorphisms are located on the same strand of the nucleic acid.

84. The method of claim 83, wherein the enzyme is a nickase.

85. The method of claim 84, wherein the nickase is selected from the group consisting of NY2A and NYS1.

10 86. A method for detecting methylated nucleotides within a nucleic acid sample, comprising:

splitting a nucleic acid sample into separate reaction vessels;

contacting nucleic acid in one reaction vessel with bisulfite;

amplifying the nucleic acid in each reaction vessel;

15 cleaving the nucleic acids in each reaction vessel to produce fragments thereof;

obtaining a mass spectrum of the resulting fragments from one reaction vessel and another mass spectrum of the resulting fragments from another reaction vessel; whereby:

20 cytosine methylation is detected by identifying a difference in signals between the mass spectra.

87. The method of claim 86, wherein:

the step of amplifying is carried out in the presence of uracil; and

the step of cleaving is effected by a uracil glycosylase.

25 88. A method for identifying a biological sample, comprising:  
generating a data set indicative of the composition of the biological sample;

denoising the data set to generate denoised data;

deleting the baseline from the denoised data to generate an intermediate

30 data  
set;

defining putative peaks for the biological sample;

-126-

- using the putative peaks to generate a residual baseline;  
removing the residual baseline from the intermediate data set to generate  
a corrected data set;  
locating, responsive to removing the residual baseline, a probable peak in  
5 the  
corrected data set; and  
identifying, using the located probable peak, the biological sample;  
wherein the generated biological sample data set comprises data from  
sense  
10 strands and antisense strands of assay fragments.
89. The method according to claim 88, wherein identifying includes  
combining  
data from the sense strands and the antisense strands, and comparing the data  
against expected sense strand and antisense strand values, to identify the  
15 biological  
sample.
90. The method according to claim 88, wherein identifying includes  
deriving a peak probability for the probable peak, in accordance with whether the  
probable peak is from sense strand data or from antisense strand data.
- 20 91. The method according to claim 88, wherein identifying includes  
deriving a peak probability for the probable peak and applying an allelic penalty in  
response to a  
ratio between a calculated area under the probable peak and a calculated  
expected average area under all peaks in the data set.
- 25 92. A method for identifying a biological sample, comprising:  
generating a data set indicative of the composition of the biological  
sample;  
denoising the data set to generate denoised data;  
deleting the baseline from the denoised data to generate an intermediate  
30 data  
set;

-127-

defining putative peaks for the biological sample; using the  
putative peaks to generate a residual baseline;  
removing the residual baseline from the intermediate data set to generate  
a

5 corrected data set;

locating, responsive to removing the residual baseline, a probable peak in  
the corrected data set; and

identifying, using the located probable peak, the biological sample;

wherein identifying includes deriving a peak probability for the probable  
10 peak and

applying an allelic penalty in response to a ratio between a calculated  
area under the  
probable peak and a calculated expected average area under all peaks in the data  
set.

15 93. The method according to claim 92, wherein identifying includes  
comparing

data from probable peaks that did not receive an applied allelic penalty to  
determine their mass in accordance with oligonucleotide biological data.

94. The method according to claim 92, wherein the allelic penalty is  
20 not applied to probable peaks whose ratio of area under the peak to the  
expected area value is greater than 30%.

95. A method for detecting a polymorphism in a nucleic acid,  
comprising:

25 amplifying a region of the nucleic acid to produce an amplicon, wherein  
the resulting amplicon comprises one or more enzyme restriction sites;

contacting the amplicon with a restriction enzyme to produce fragments;  
obtaining a mass spectrum of the resulting fragments and analyzing  
signals in the mass spectrum by the method of claim 88; whereby:

the polymorphism is detected from the pattern of the signals.

30 96. A subcollection of samples from a target population, comprising:  
a plurality of samples, wherein the samples are selected from the group  
consisting of nucleic acids, fetal tissue, protein samples; and

-128-

a symbology on the containers containing the samples, wherein the symbology is representative of the source and/or history of each sample, wherein:

5 the target population is a healthy population that has not been selected for any disease state;

the collection comprises samples from the healthy population; and

the subcollection is obtained by sorting the collection according to specified parameters.

97. The combination of claim 26, wherein the samples are selected  
10 selected from the group consisting of nucleic acids, fetal tissue, protein, tissue, body fluid, cell, seed, microbe, pathogen and reproductive tissue samples.

98. A combination, comprising the database of claim 8 and a mass spectrometer.

99. The combination of claim 98 that is an automated process line for  
15 analyzing biological samples.

100. A system for high throughput processing of biological samples, comprising:

a database of claim 8, wherein the samples tested by the automated  
process line comprise samples from subjects in the database; and  
20 a mass spectrometry for analysis of biopolymers in the samples.

## DNA Bank

Number of Samples	3912
-------------------	------

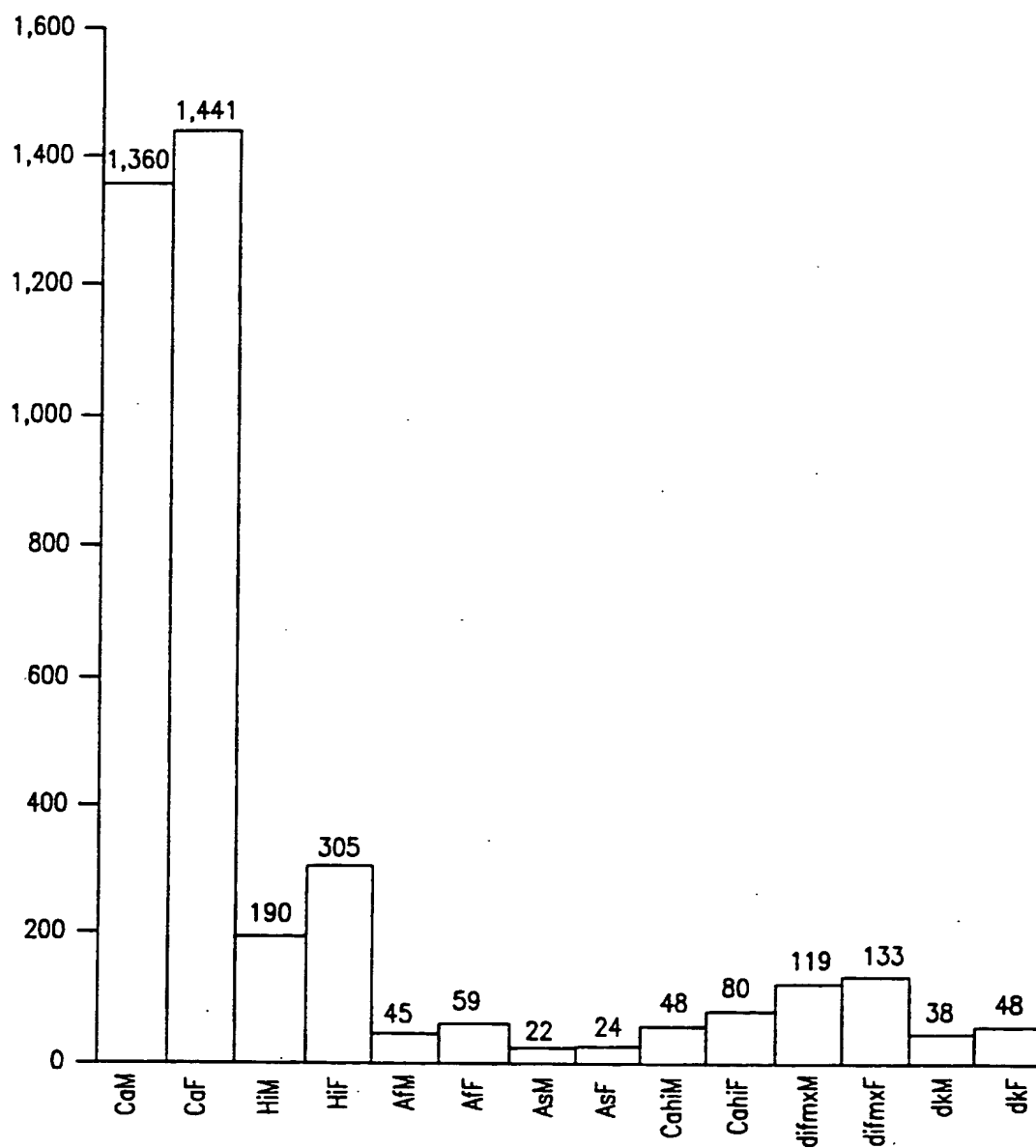


FIG. 1A

2 / 51

Caucasians

Number of Samples	2801
-------------------	------

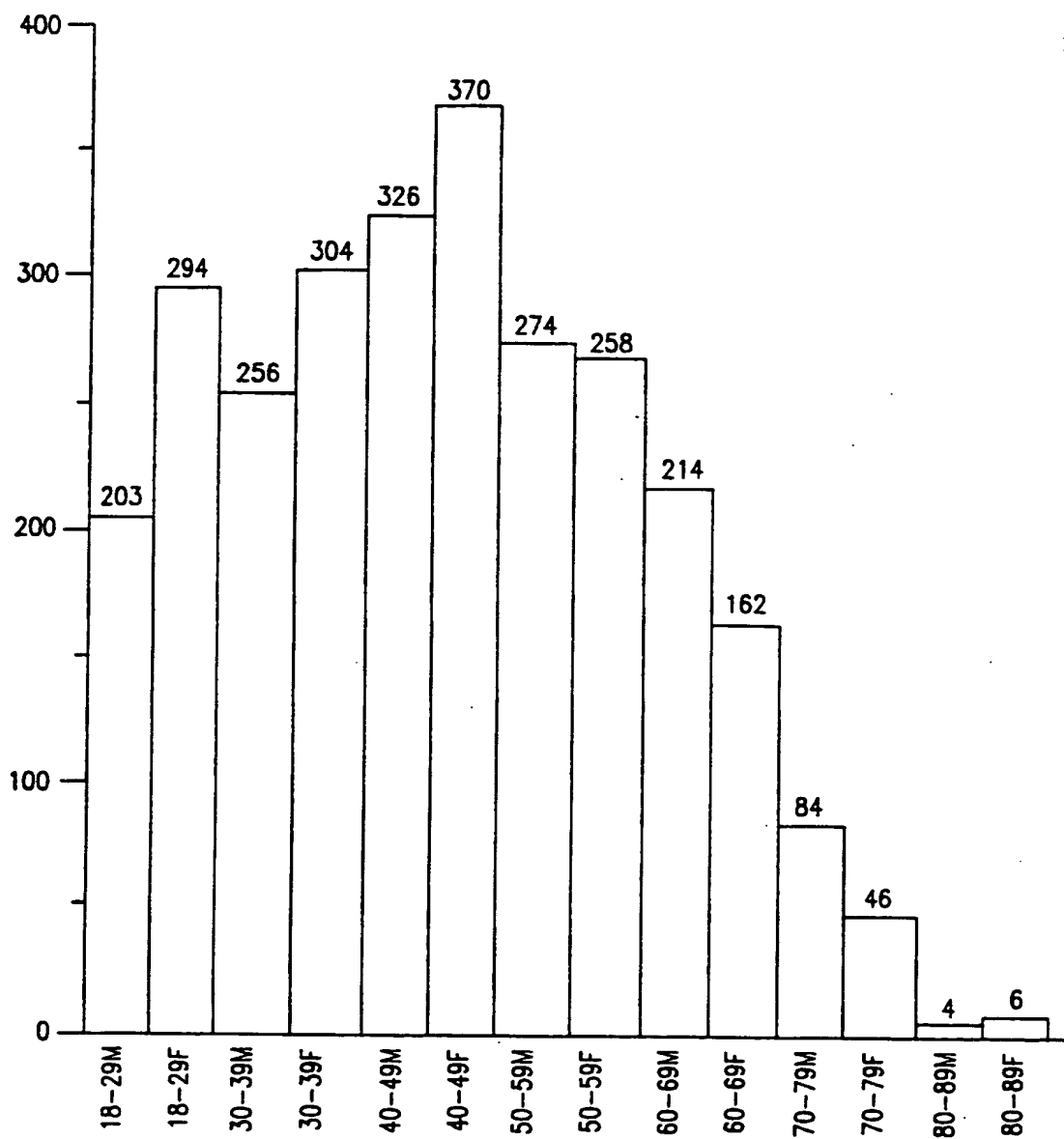


FIG. 1B

3/51

Hispanics

Number of Samples	495
-------------------	-----

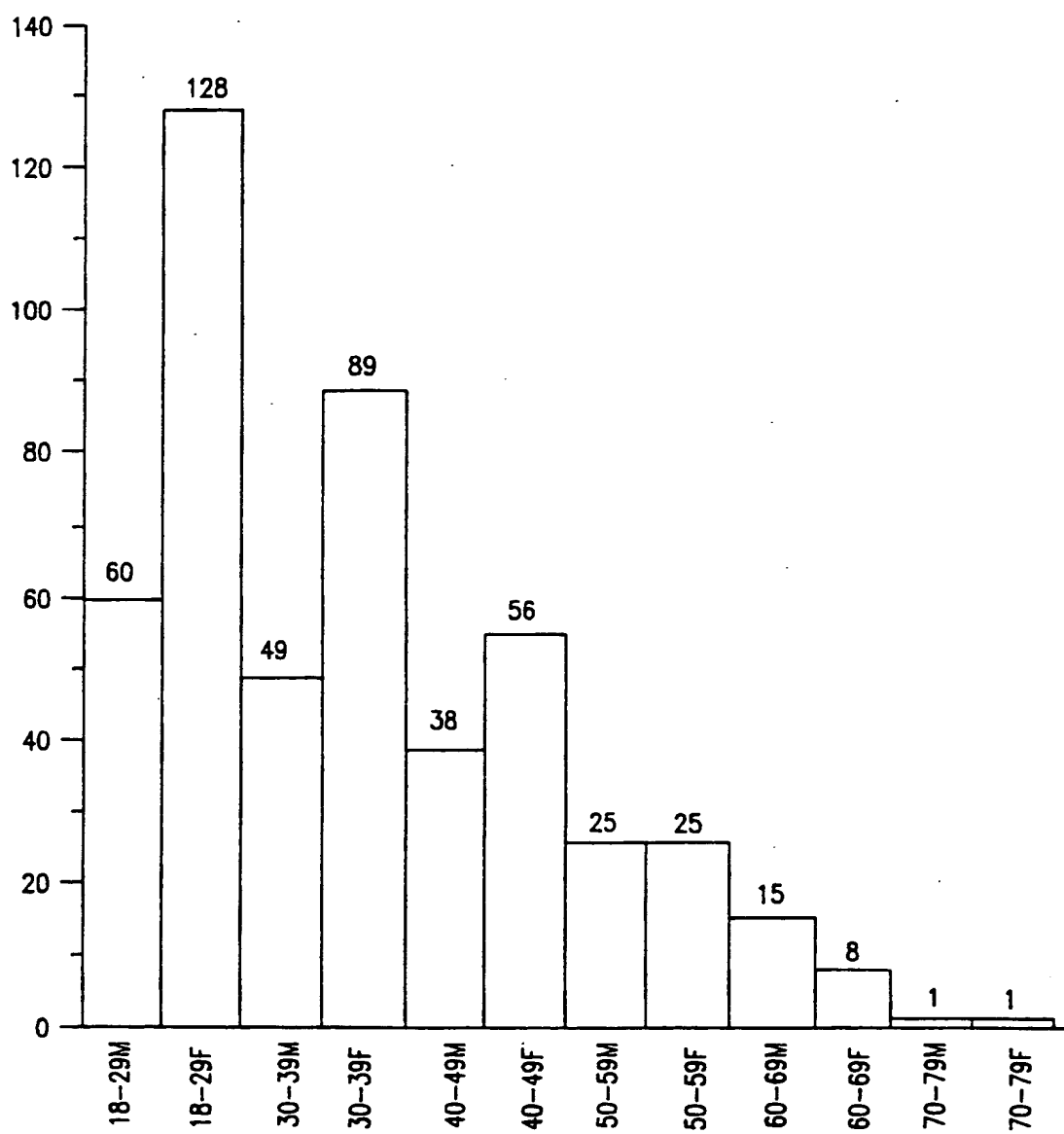
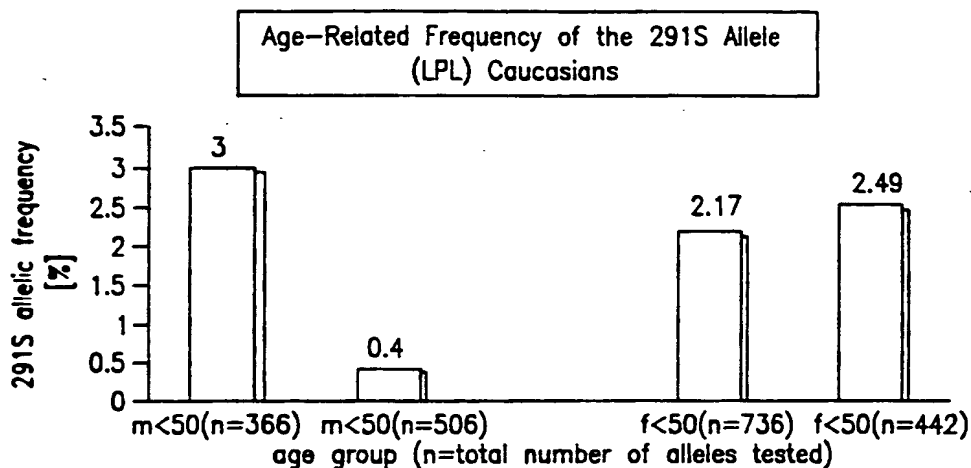
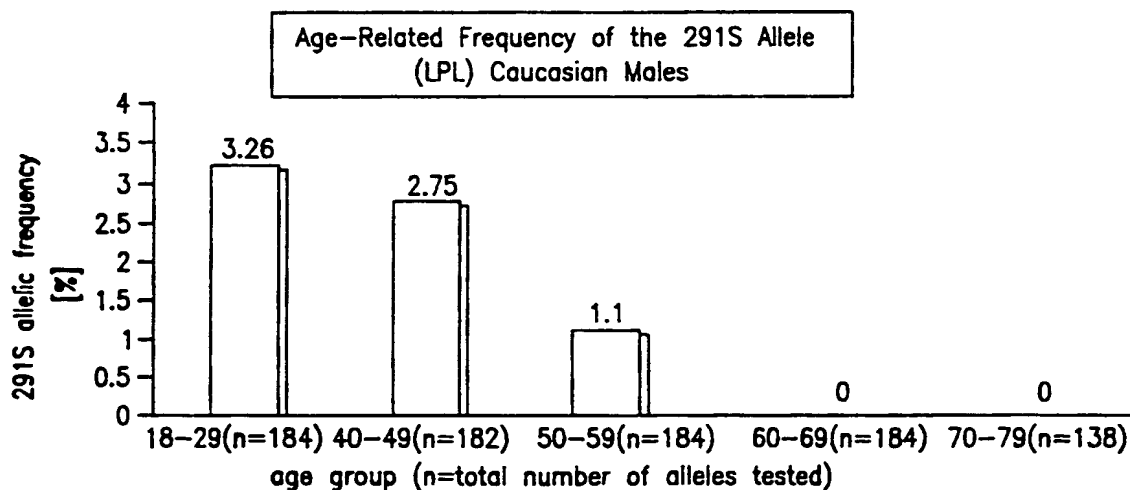
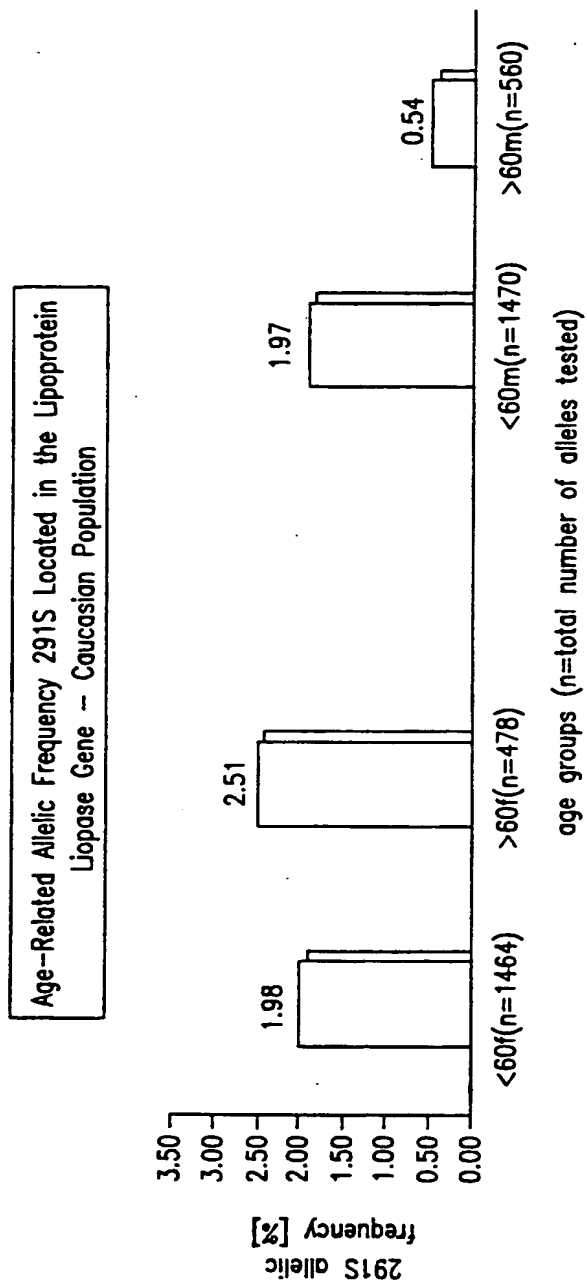


FIG. 1C

4/51

**FIG. 2A****FIG. 2B**

5/ 51



Significance: Allele frequency of Ser291 drops from 1.97% to 0.54% in males;  $p=0.009$

FIG. 2C

6/51

Questionnaire for  
Population-Based  
Sample Banking

Data Collection Form

Collection Information

Consent Form Signed Yes No

Date of Collection (MM/DD/YY)\_\_\_/\_\_\_/98

Time of Sample Collection(nearest hour in 24 hour clock format)\_\_\_\_\_

Initials of Data Collector\_\_\_\_\_Collecting Agency\_\_\_\_\_

(DO NOT COMPLETE: (For Data Entry Only)Sample\_\_\_\_\_intact\_\_\_\_\_lost \_\_\_\_\_broken

Affix Barcode Here

Donor information

Sex: ☐ Male ☐ Female

Date of Birth (MM/YY)\_\_\_/\_\_\_

In which state do you live? \_\_\_\_\_ How long have you lived there ? \_\_\_\_\_ Years

What is your highest grade you completed in school?

☐ less than 8th grade☐ 8th,9th,10th or 11th grade☐ high school graduate or equivalency☐ some college 2 yr degree☐ college graduate 4 yr degree☐ post graduate education or degree

To the best of your knowledge what is the Ethnic Origin of your:

Father

Mother

☐☐

Caucasian (please check specific geographic area below if known)

☐☐

Northern Europe (Austria,Denmark,Finland,France,Germany,Netherlands,Norway,Sweden,Switzerland,U.K.)

☐☐

Southern Europe (Greece,Italy,Spain)

☐☐

Eastern Europe (Czechoslovakia,Hungary,Poland,Russia,Yugoslavia)

☐☐

Middle Eastern (Israel,Egypt,Iran,Iraq,Jordan,Syria, other Arab States)

☐☐

African-American

☐☐

Hispanic (please check specific geographic area below if known)

☐☐

Mexico

☐☐

Central America,South American

☐☐

Cuba,Puerto Rico, other Caribbean

☐☐

Asian (please check specific geographic area below if known)

☐☐

Japanese

☐☐

Chinese

☐☐

Korean

☐☐

Vietnamese

☐☐

other Asian

☐☐

Other \_\_\_\_\_

☐☐

Don't know

Health information: Have you or has anyone in your immediate family(parents,brothers,sisters, or your children) had the following? Check all that apply

Disease:

You

Mother

Father

Sister

Brother

Child

Heart Disease Stroke or Arteriosclerosis

Cancer (Specify type if known)

Alzheimer's Disease or Dementia

Chronic inflammatory or Autoimmune Disease

Nervous System Disease like Multiple Sclerosis

Other (please specify)

Additional health information details you would like to provide:

FIG. 3

SUBSTITUTE SHEET (RULE 26)

7/51

# Sample Banks

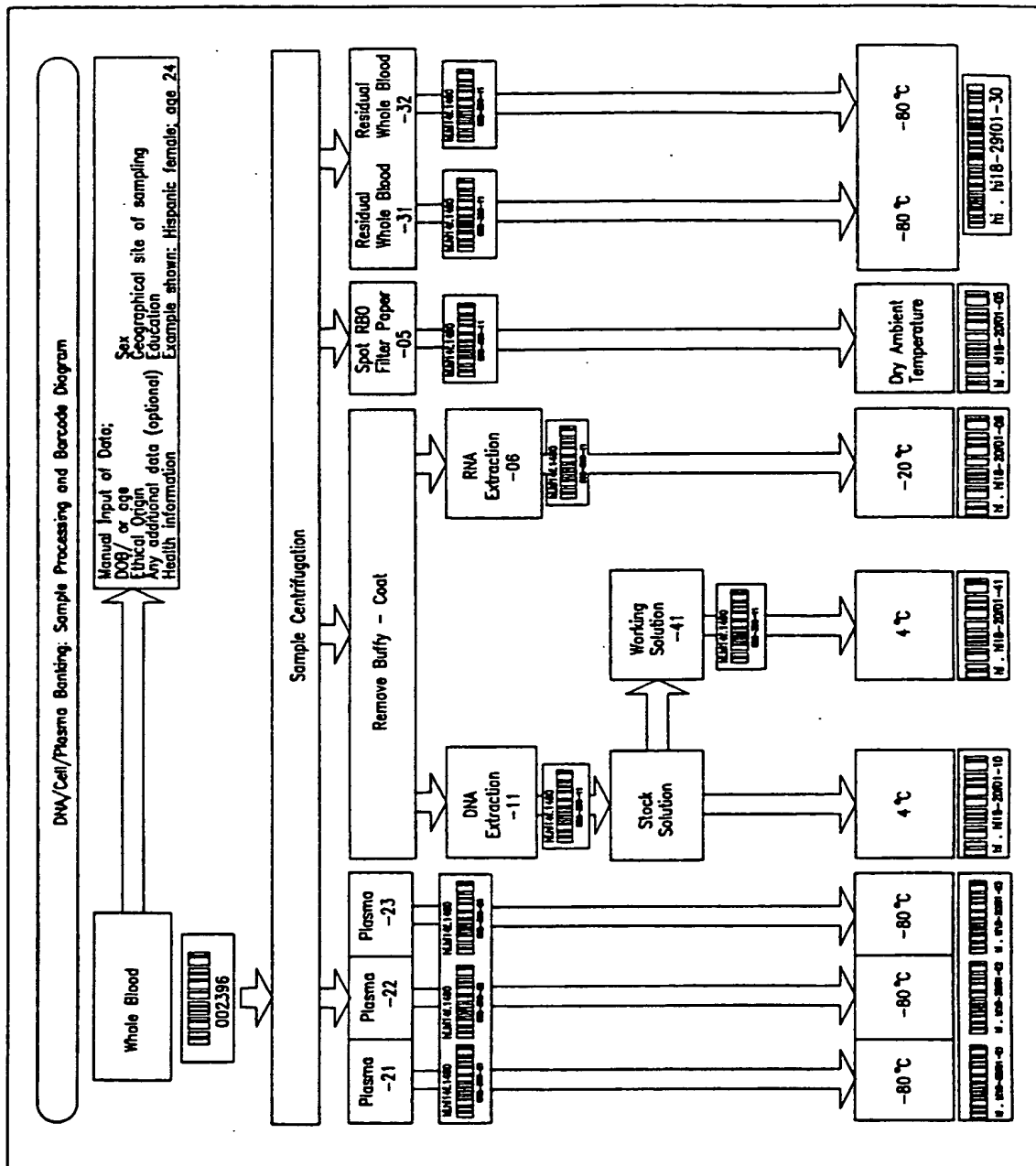


FIG. 4

8/51

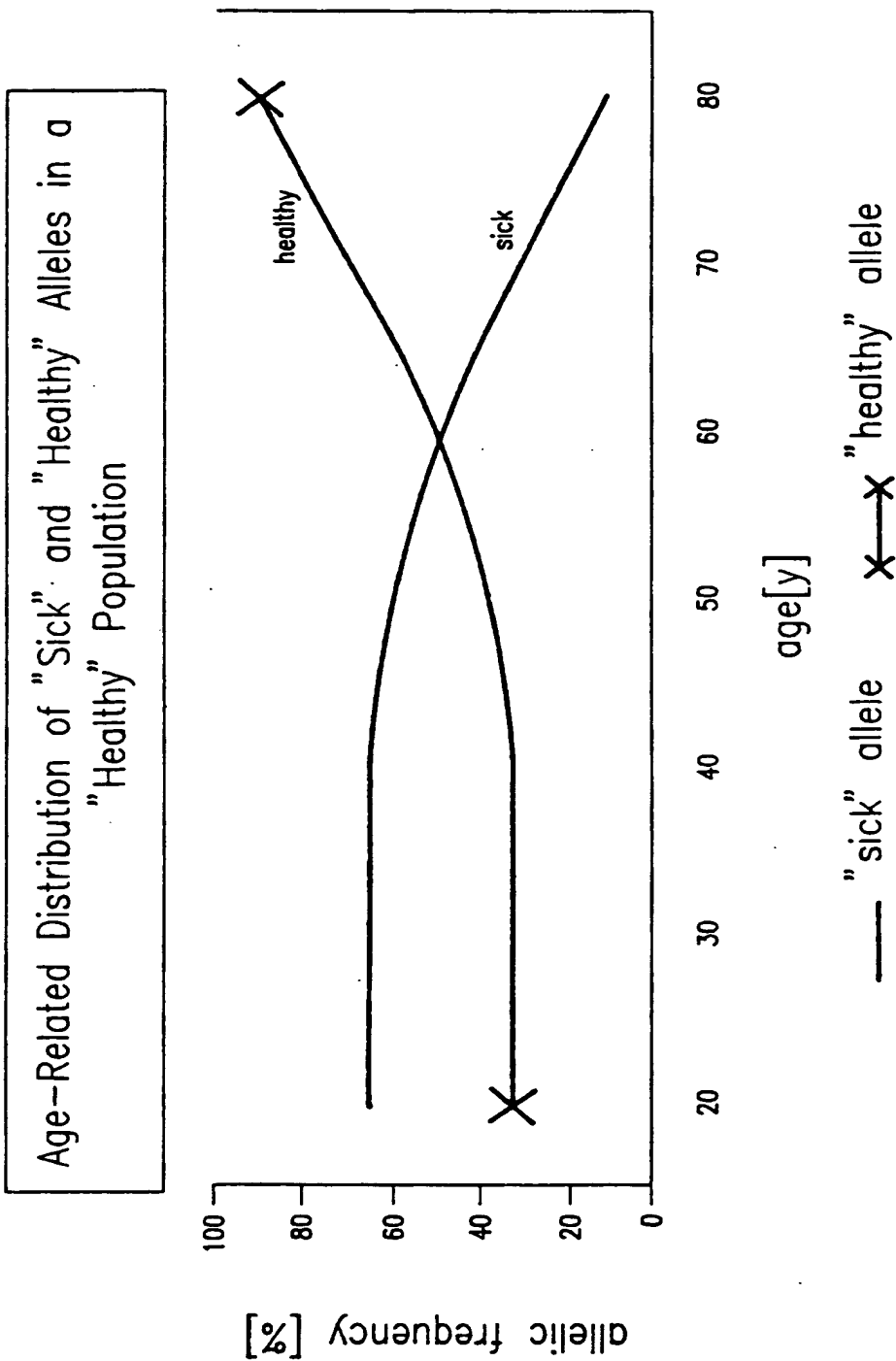


FIG. 5

9/ 51

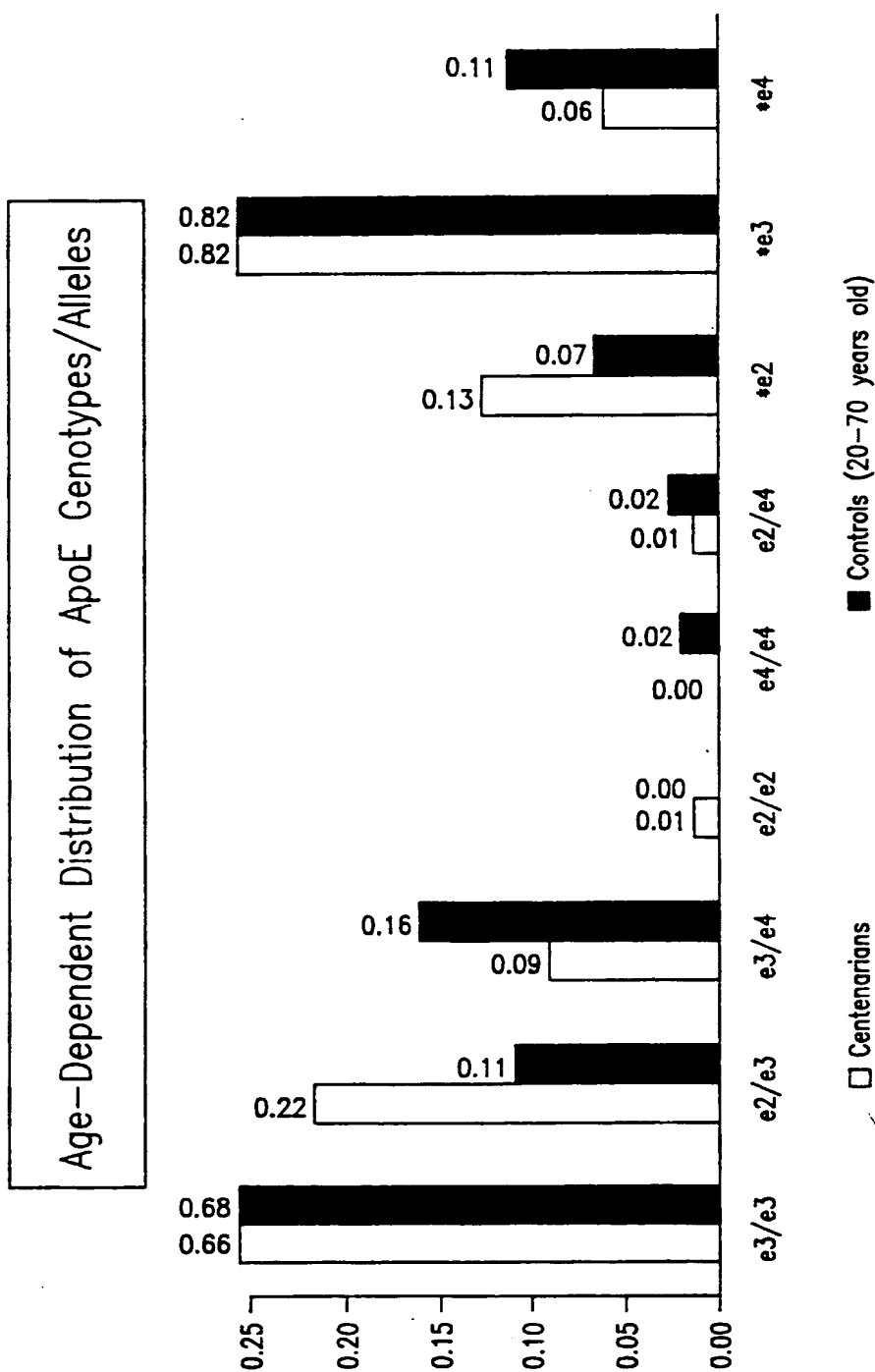


FIG. 6

10/51

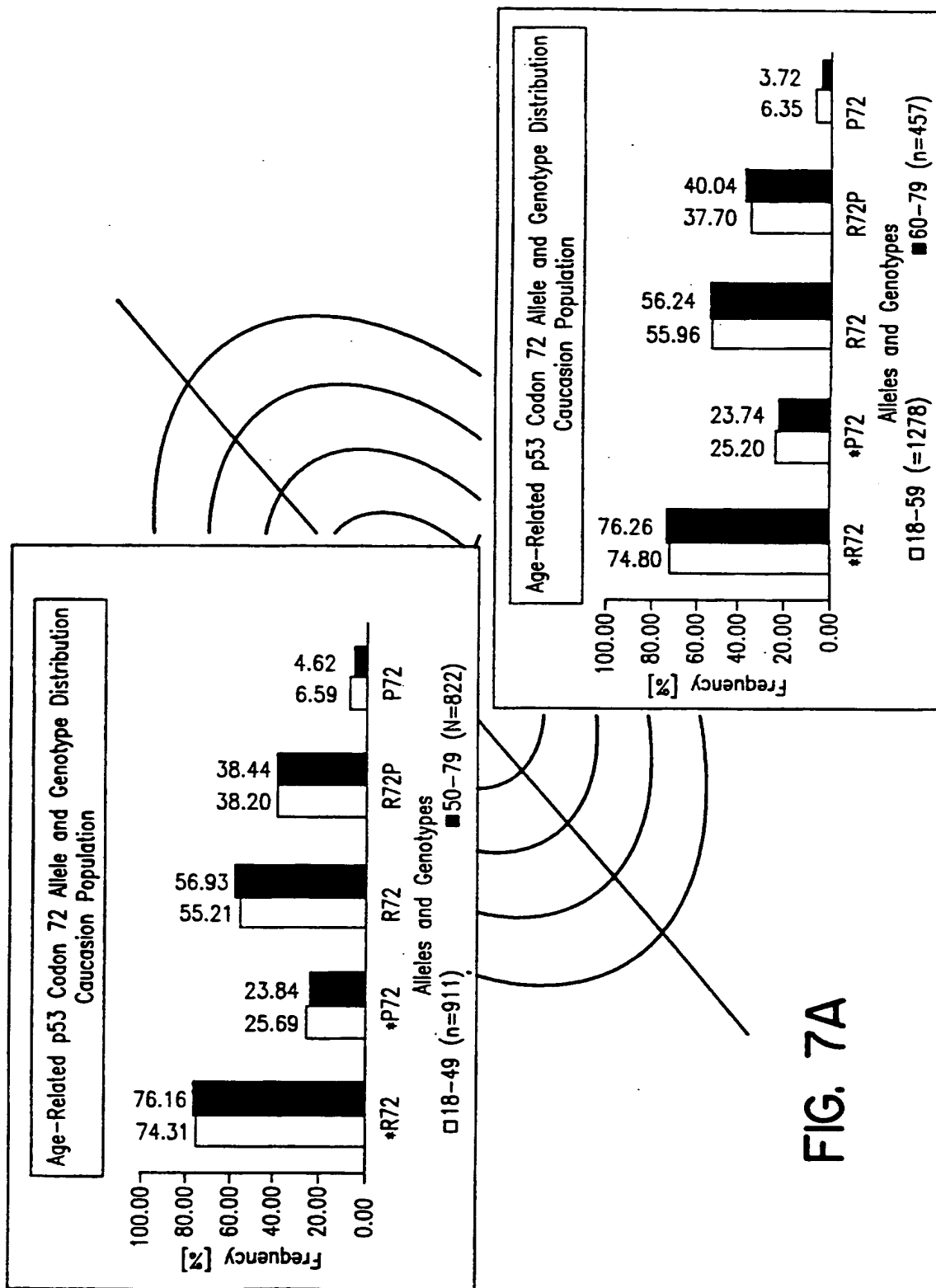


FIG. 7A

11/51

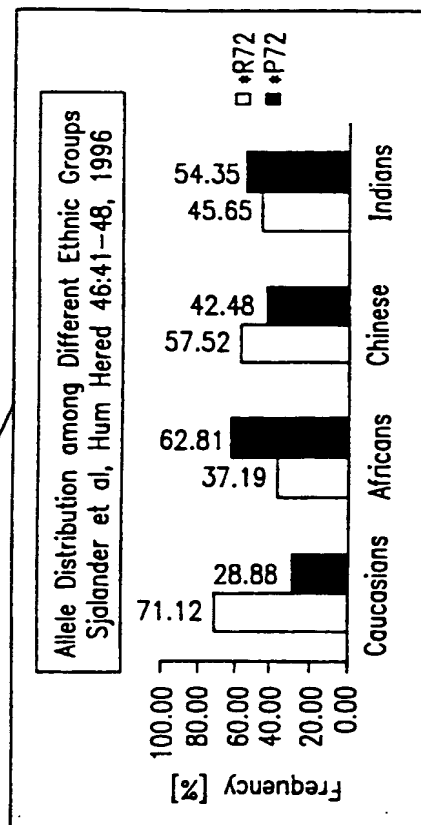
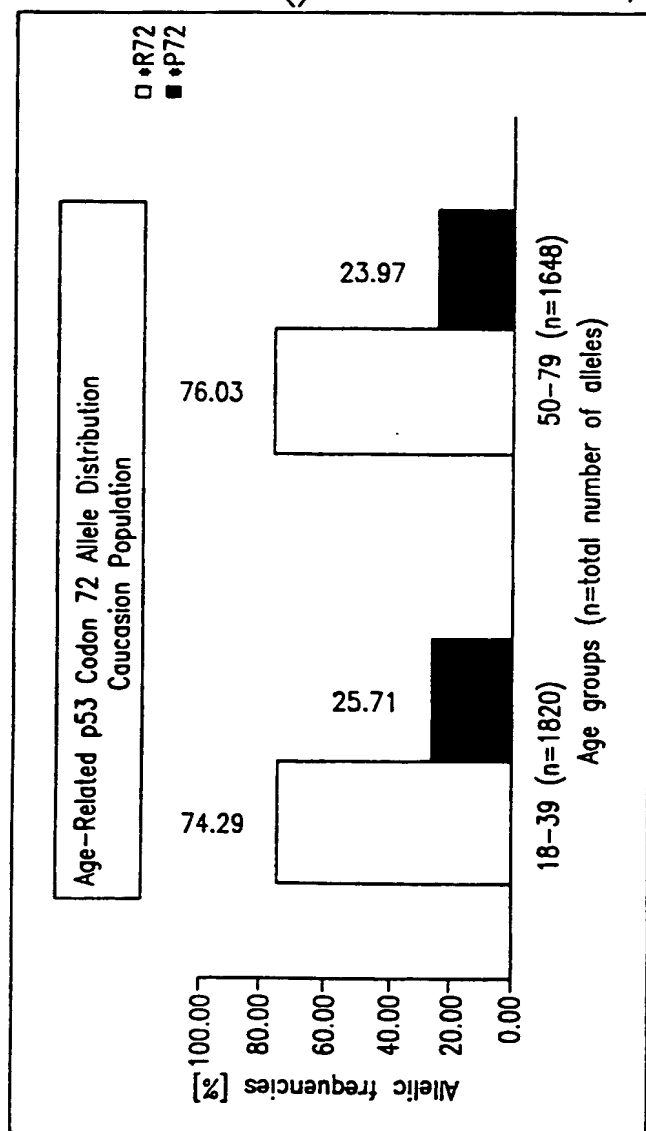
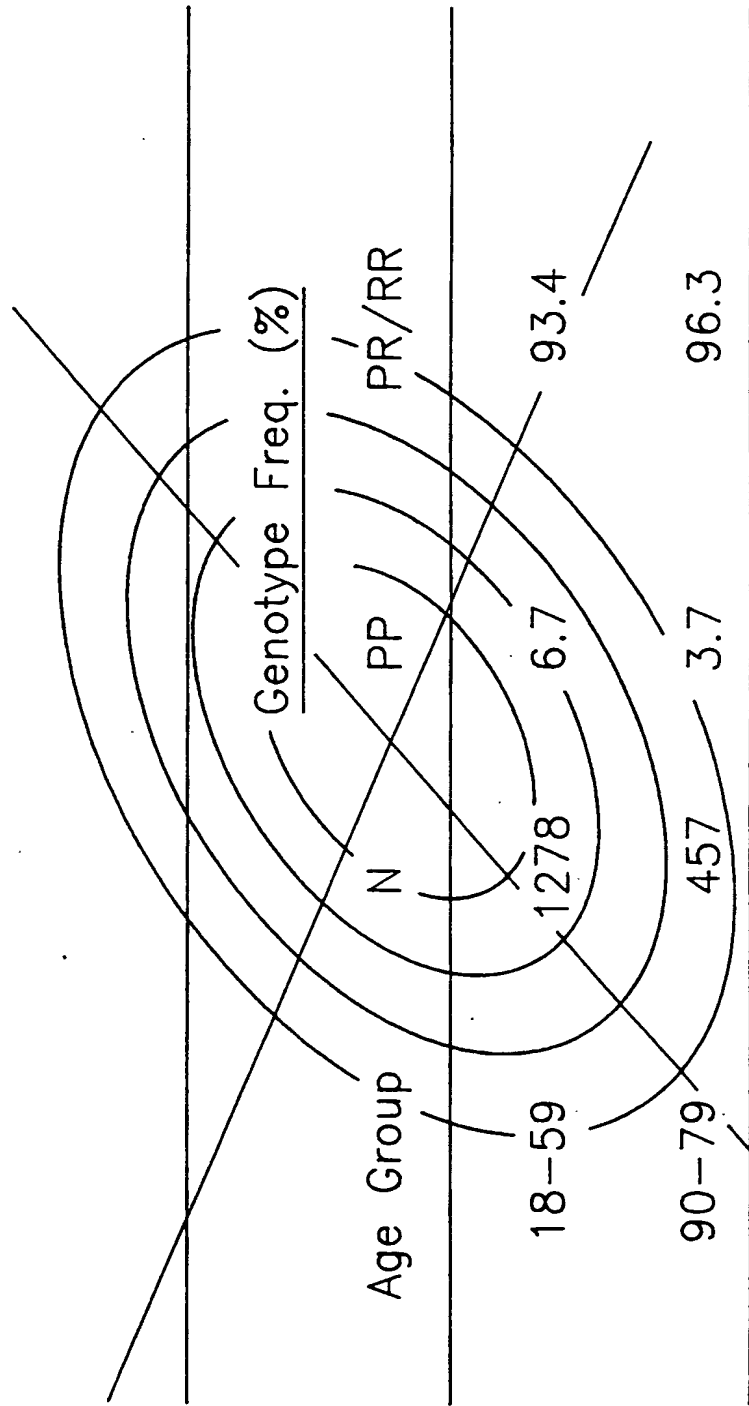


FIG. 7B

12/51

**FIG. 7C**  
P53 PP vs. PR/RR Genotype Distribution  
By Age cut point = 59



Sample Size : 1735  
 $\chi^2$ : 5.2 (1 d.f.),  $P = 0.02$

13/51

Genomic Organization of the p53 Gene

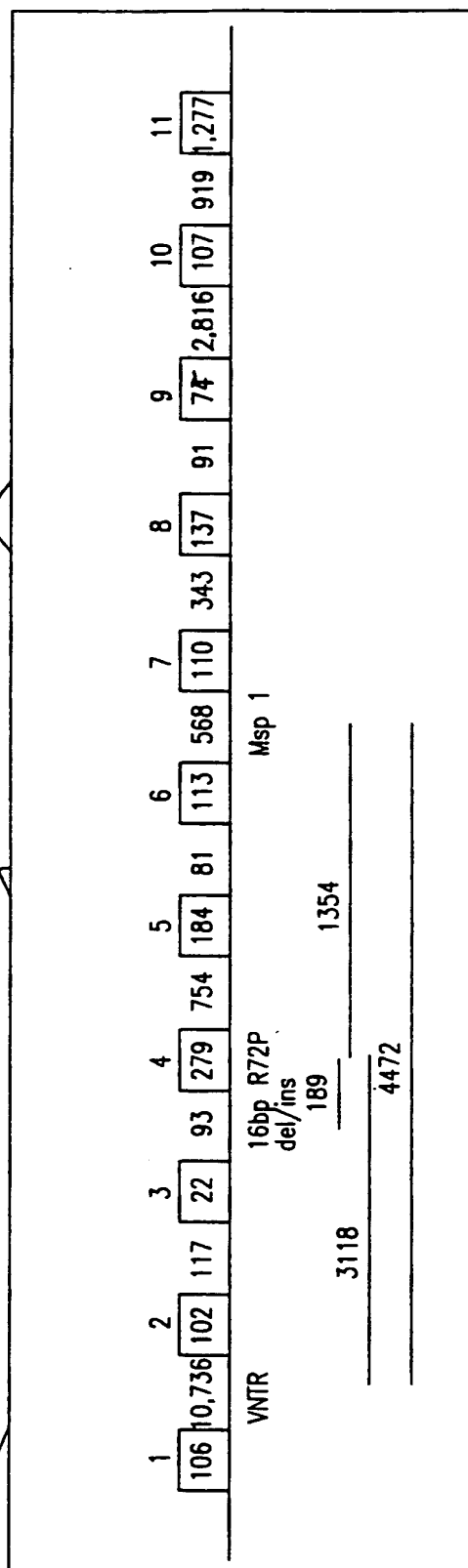


FIG. 7D

14/51

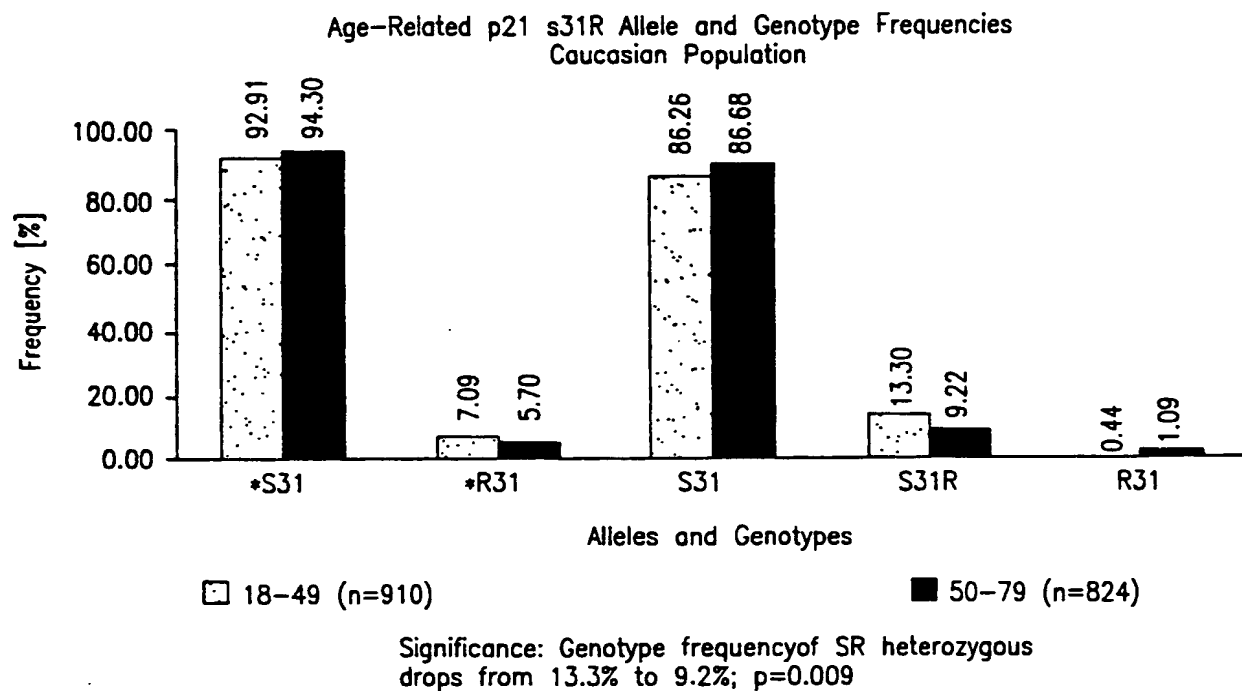


FIG. 8

15/51

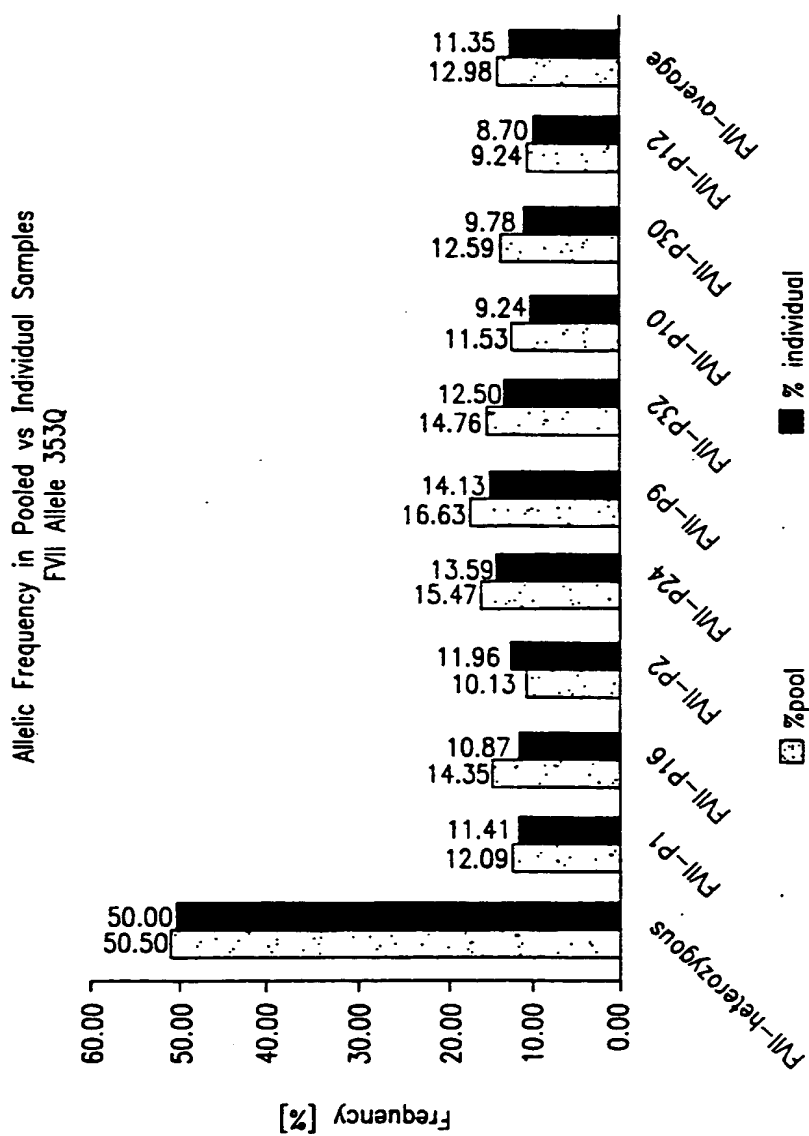


FIG. 9

16/51

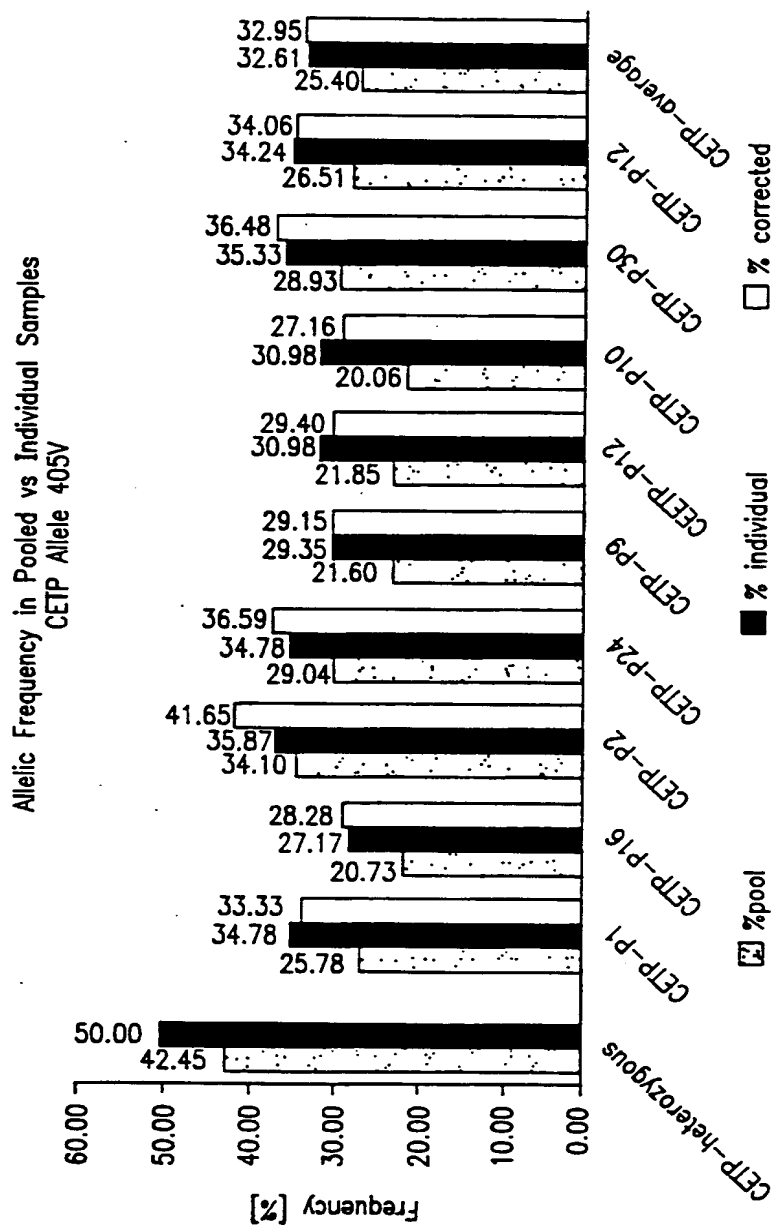


FIG. 10

17/51

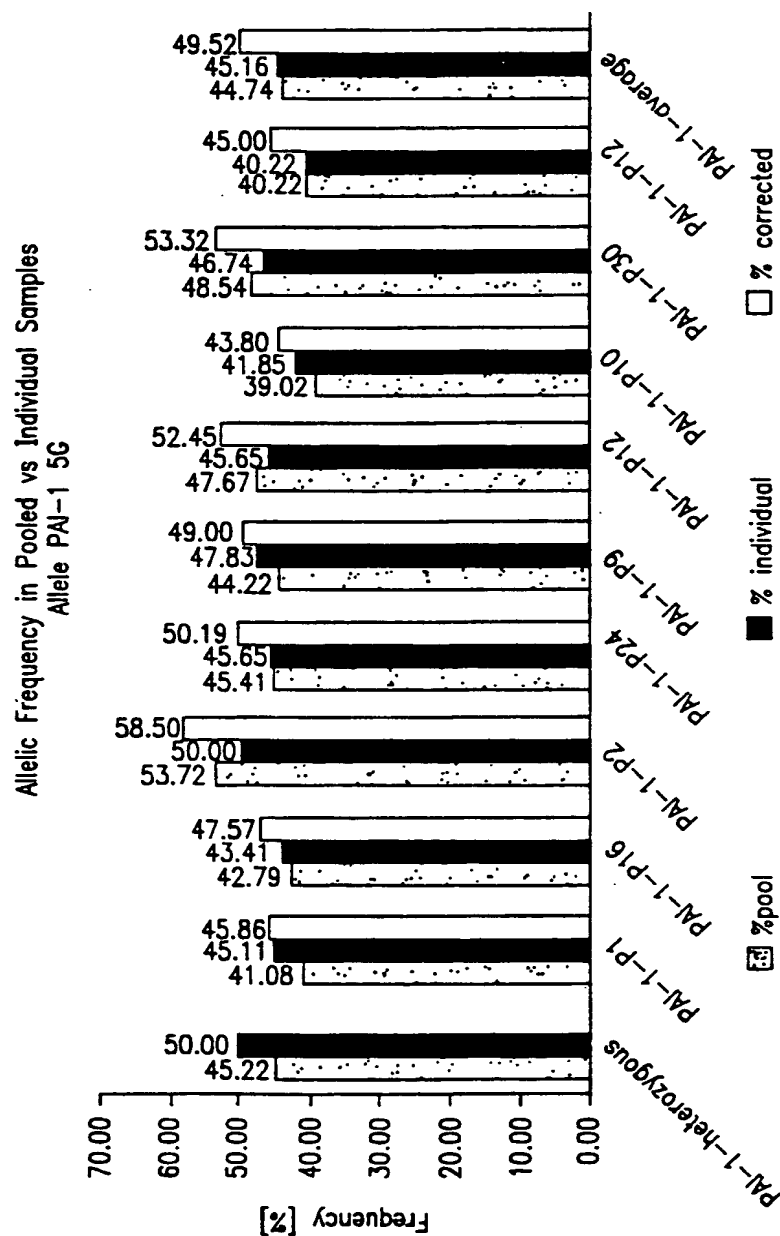
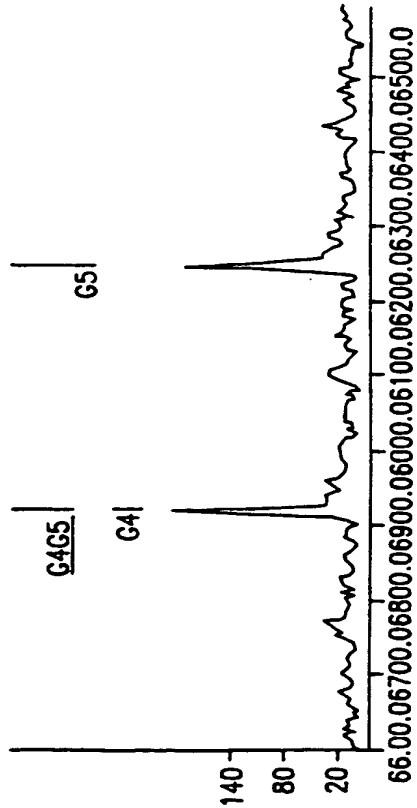
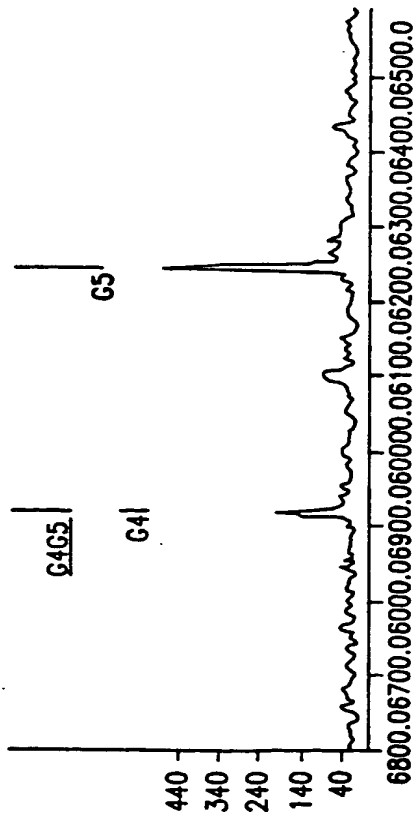
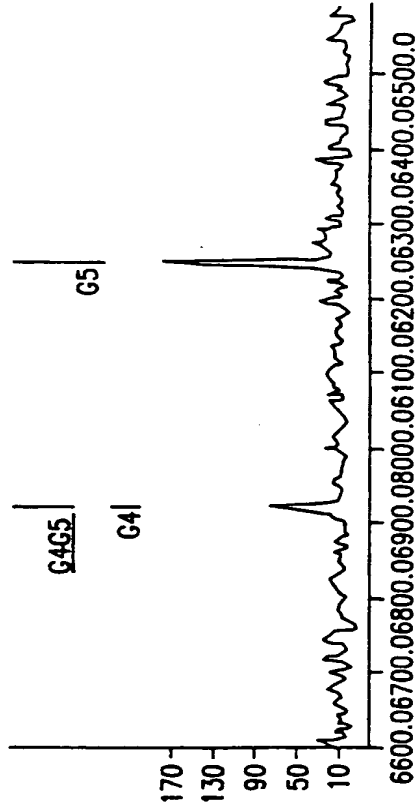
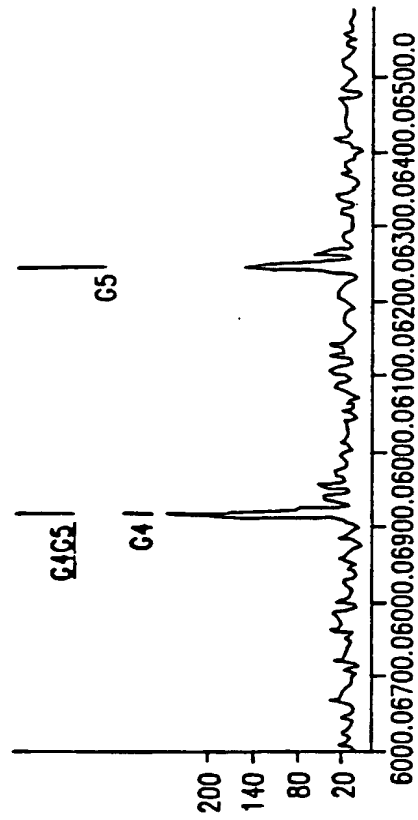
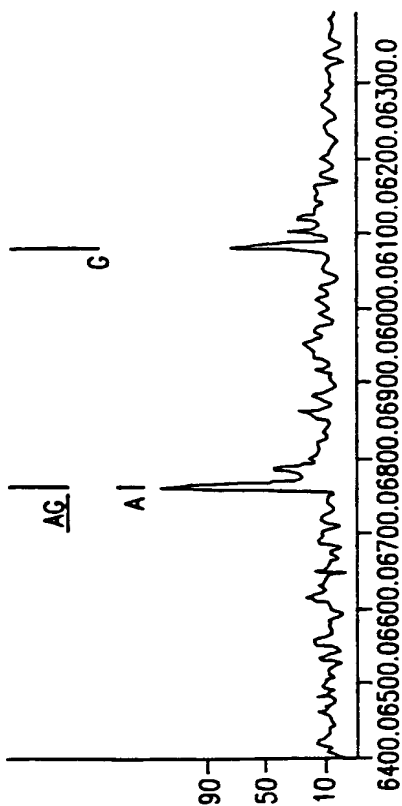


FIG. 11

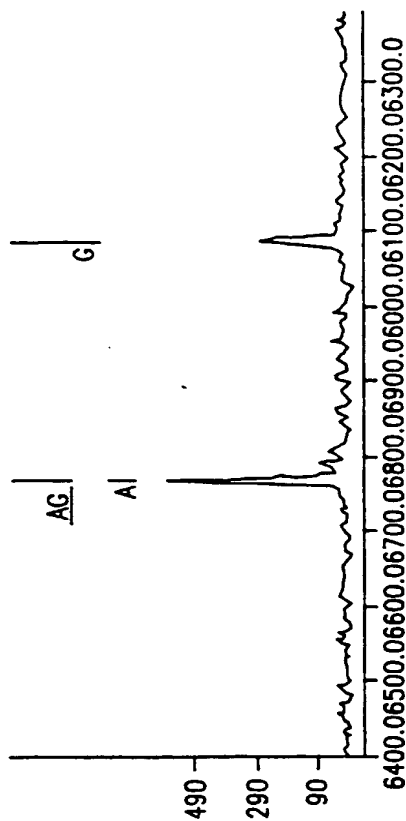


19/51



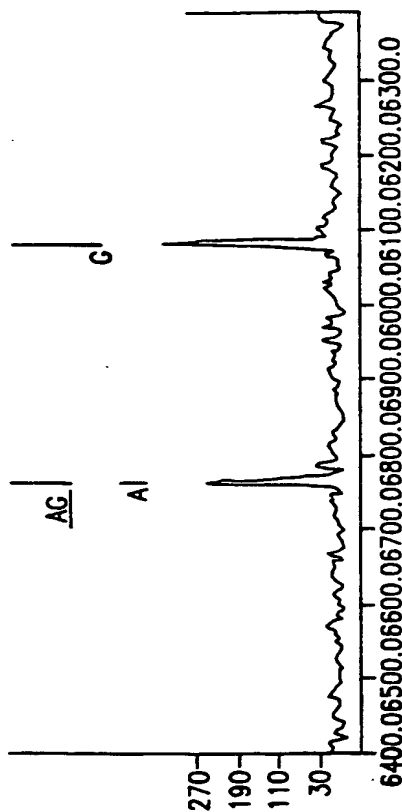
Asian Ile 60.7% Val 39.3%

FIG. 13B



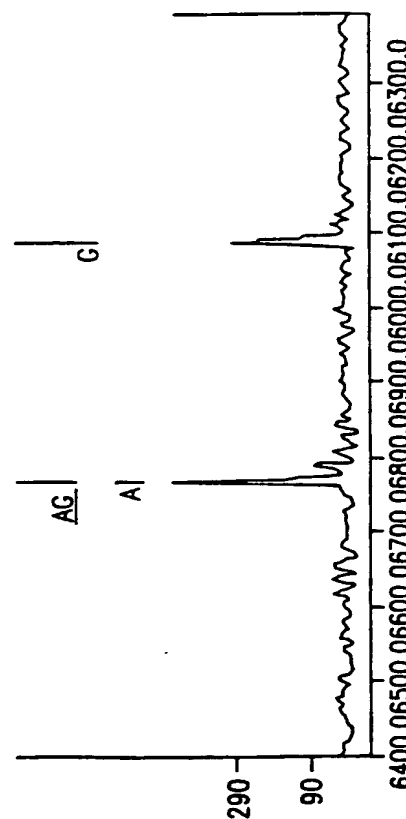
Hispanic Ile 70.1% Val 29.9%

FIG. 13D



African Ile 43.6% Val 56.4%

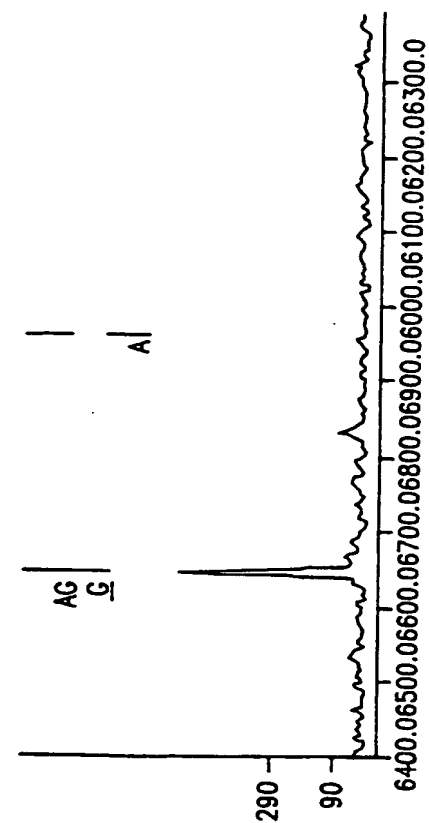
FIG. 13A



Caucasian Ile 63.0% Val 37.0%

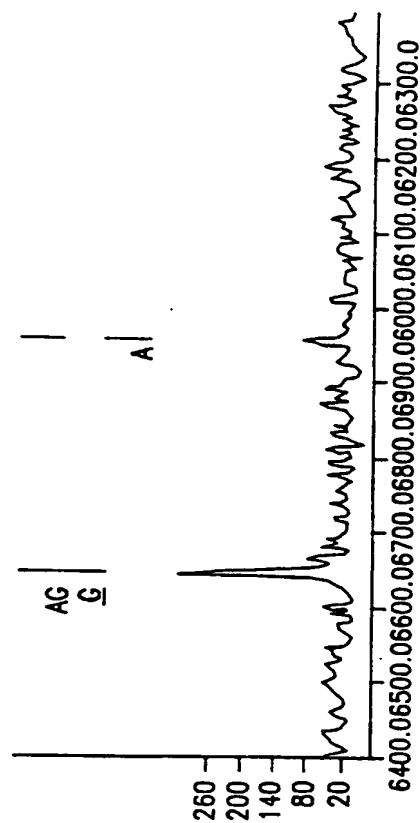
FIG. 13C

20/51



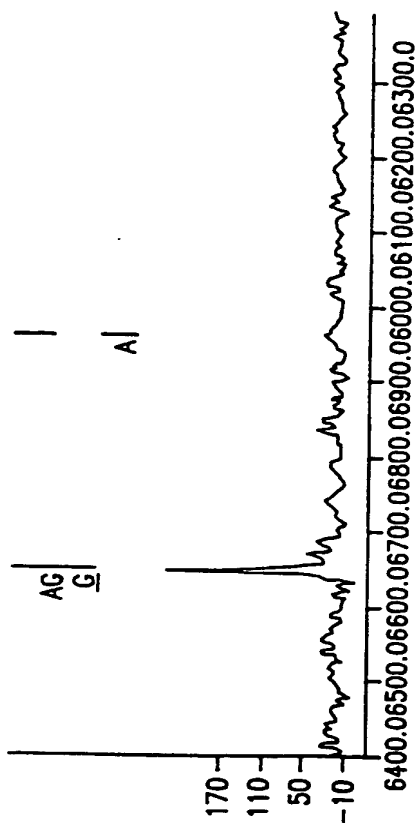
Asian Arg>95% Glu<5%

FIG. 14B



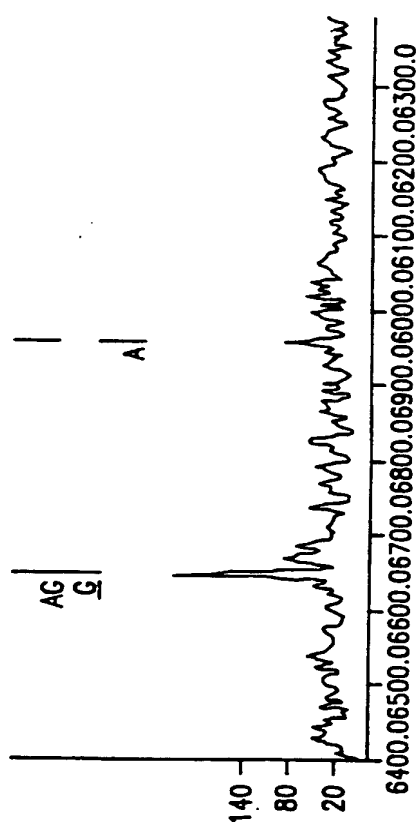
Hispanic Arg 89.2% Glu 10.8%

FIG. 14D



African Arg>95% Glu<5%

FIG. 14A



Caucasian Arg 88.5% Glu 11.5%

FIG. 14C

21/51

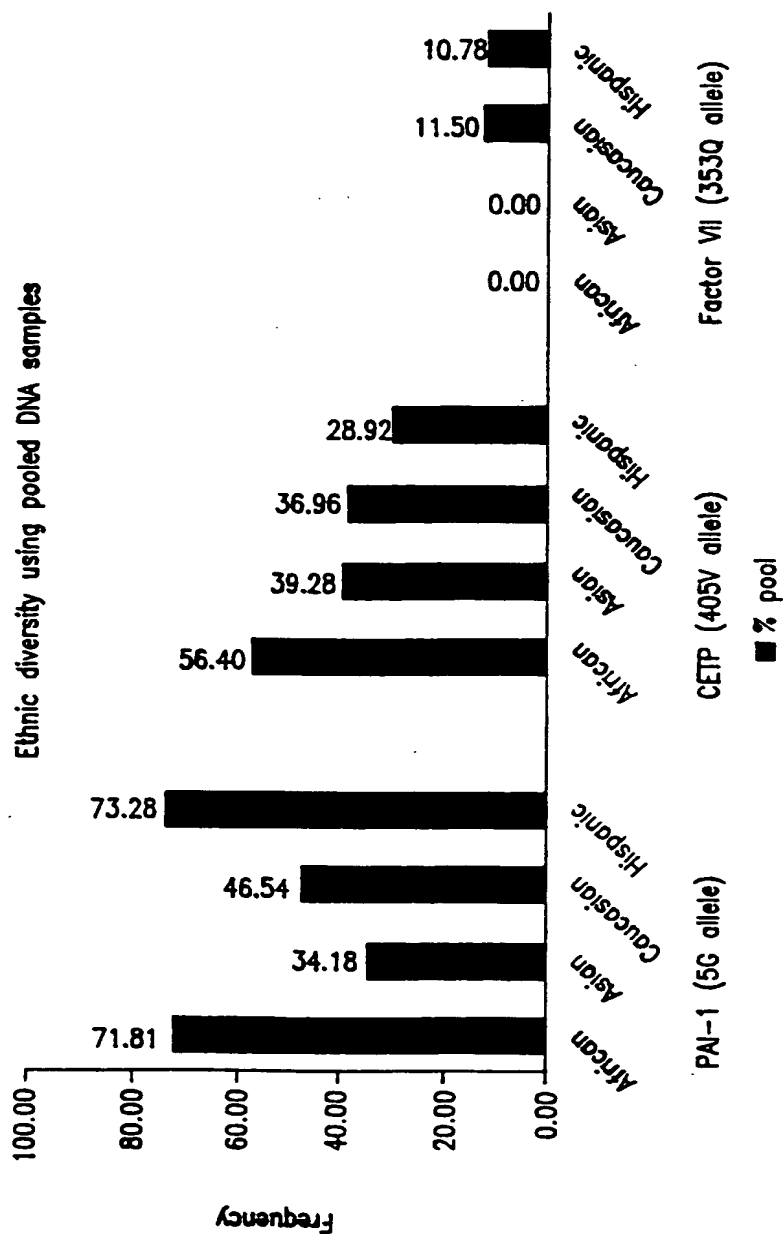
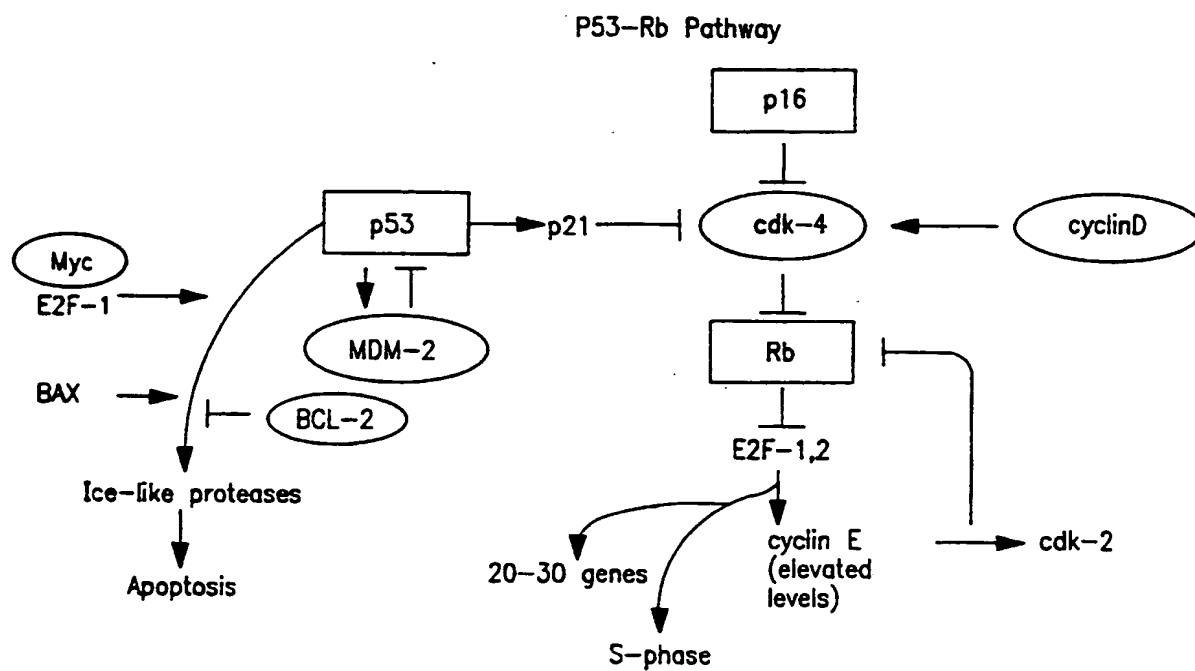


FIG. 15

22/ 51

**FIG. 16**

23/51

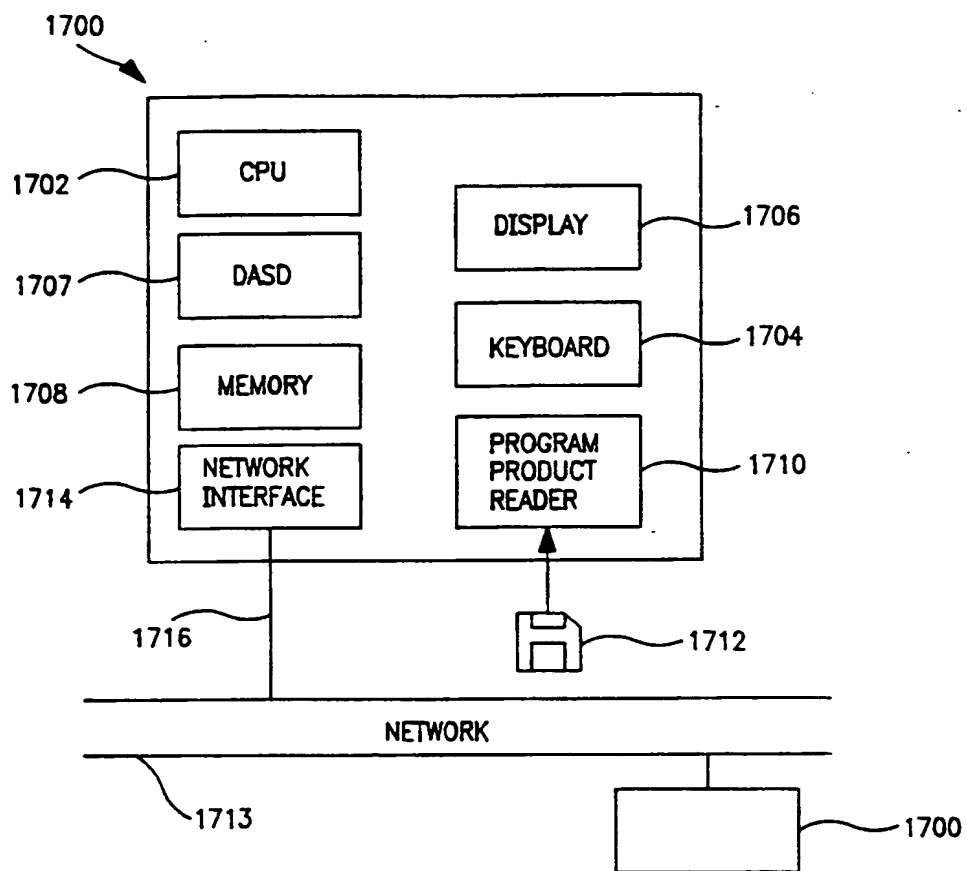


FIG. 17

24/51

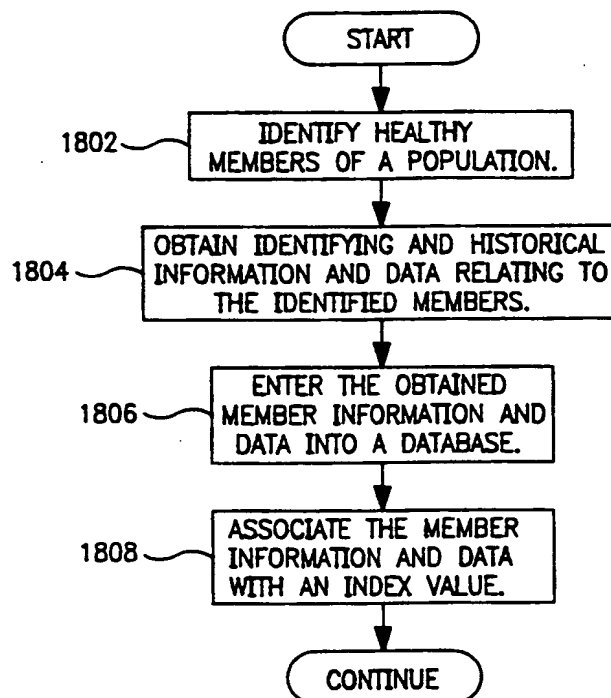


FIG. 18

25/51

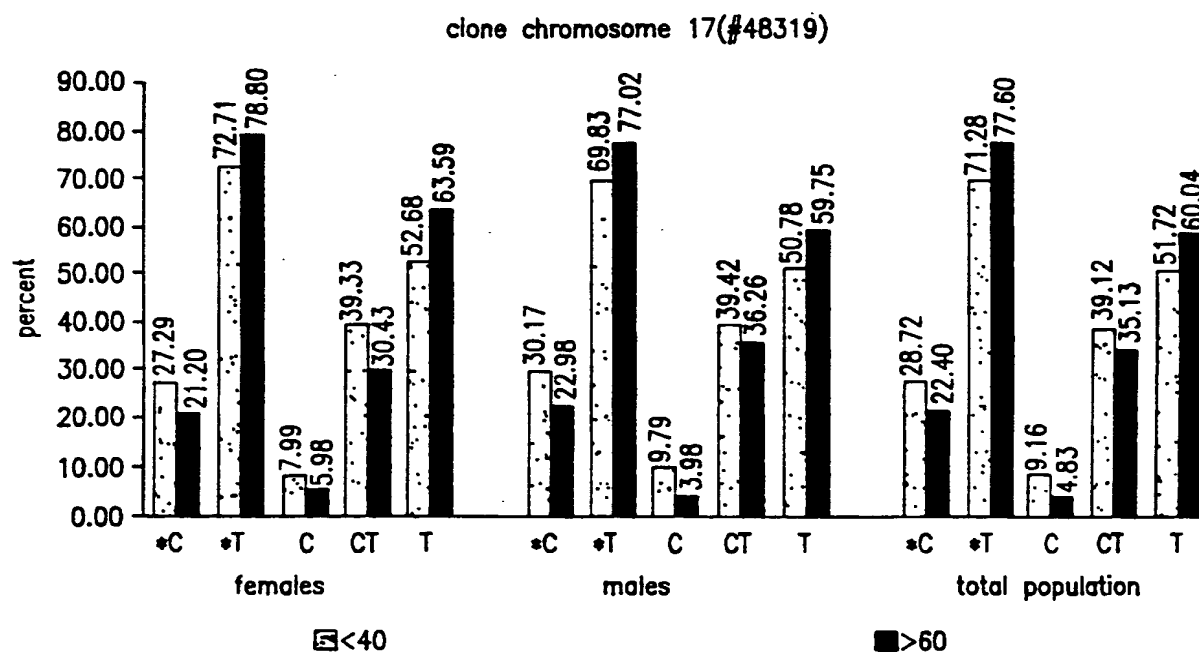


FIG. 19

26/51

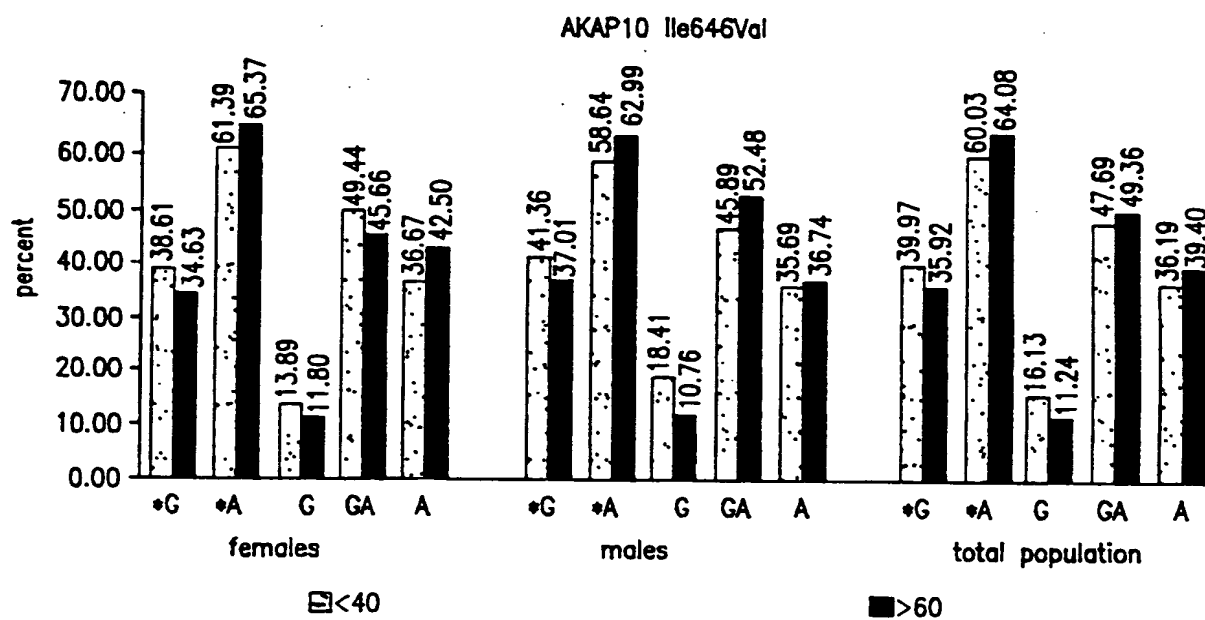


FIG. 20

27/51

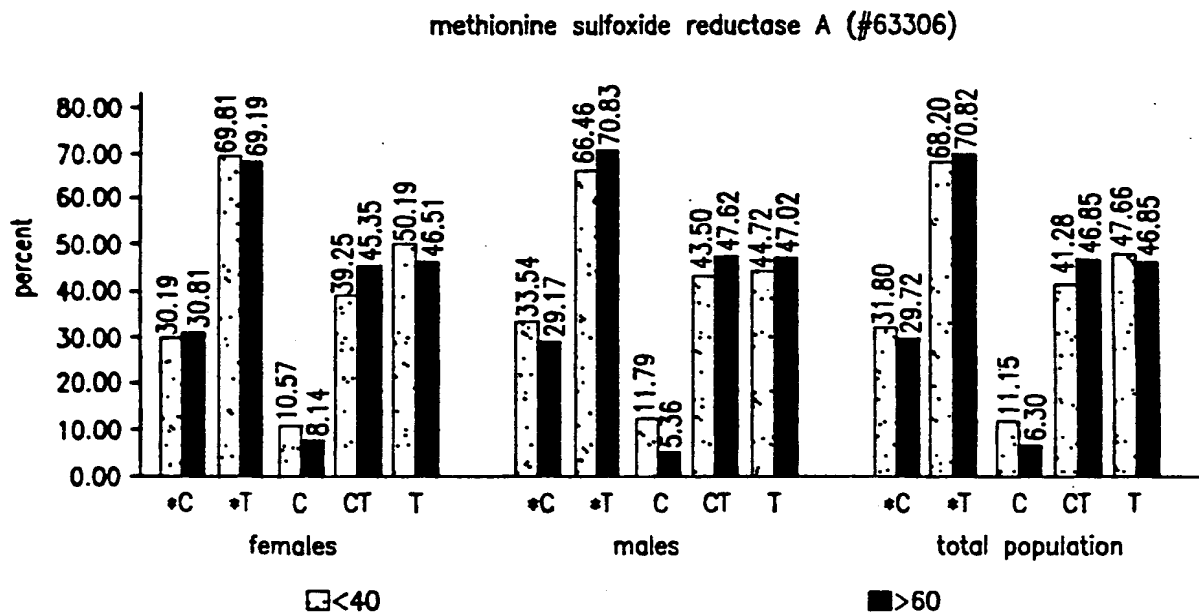


FIG. 21

28/ 51

<b>Collection Information</b>																																																																								
Consent Form Signed <input type="checkbox"/> Yes <input type="checkbox"/> No					Time of Sample Collection (nearest hour, in 24 hour clock format)		Initials		Initials of Data Collector																																																															
Date of Collection Month   Day   Year JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC									(DO NOT COMPLETE: for data entry only)  Sample: <input type="checkbox"/> Intact <input type="checkbox"/> Lost <input type="checkbox"/> Broken           Volume (ml)																																																															
<b>Donor Information</b>																																																																								
Date of Birth Month   Year JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC					Sex: <input type="checkbox"/> Male <input type="checkbox"/> Female		Height Ft. Inches		Weight (lb)		What Physical activity do you do on a regular basis? <input type="checkbox"/> Running <input type="checkbox"/> Swimming <input type="checkbox"/> Biking <input type="checkbox"/> Gymnastics <input type="checkbox"/> Other _____ <input type="checkbox"/> None																																																													
											Are you a vegetarian? <input type="checkbox"/> Yes <input type="checkbox"/> No																																																													
											If Female: How many times have you been pregnant?   How many times did you give birth?																																																													
To the best of your knowledge, what is the Ethnic Origin of you:																																																																								
<table border="0" style="width: 100%;"> <tr> <td style="vertical-align: top; width: 15%;">Father</td> <td style="vertical-align: top; width: 15%;">Mother</td> <td></td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Caucasian (please mark specific geographic area below if known)</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Northern Europe (Austria, Denmark, Finland, France, Germany, Netherlands, Norway, Sweden, Switzerland, UK)</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Southern Europe (Greece, Italy, Spain, Turkey)</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Eastern Europe (Czechoslovakia, Hungary, Poland, Russia, Yugoslavia)</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Middle Eastern (Israel, Egypt, Iran, Iraq, Jordan, Syria, Other Arab States)</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>African-American</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Hispanic (please mark specific geographic area below if known)</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Mexico</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Central America, South America</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Cuba, Puerto Rico, other Caribbean</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Asian (please mark specific geographic area below if known)</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Japanese</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Chinese</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Korean</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Vietnamese</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Filipino</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Native American</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Other _____</td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td>Don't know</td> </tr> </table>													Father	Mother		<input type="checkbox"/>	<input type="checkbox"/>	Caucasian (please mark specific geographic area below if known)	<input type="checkbox"/>	<input type="checkbox"/>	Northern Europe (Austria, Denmark, Finland, France, Germany, Netherlands, Norway, Sweden, Switzerland, UK)	<input type="checkbox"/>	<input type="checkbox"/>	Southern Europe (Greece, Italy, Spain, Turkey)	<input type="checkbox"/>	<input type="checkbox"/>	Eastern Europe (Czechoslovakia, Hungary, Poland, Russia, Yugoslavia)	<input type="checkbox"/>	<input type="checkbox"/>	Middle Eastern (Israel, Egypt, Iran, Iraq, Jordan, Syria, Other Arab States)	<input type="checkbox"/>	<input type="checkbox"/>	African-American	<input type="checkbox"/>	<input type="checkbox"/>	Hispanic (please mark specific geographic area below if known)	<input type="checkbox"/>	<input type="checkbox"/>	Mexico	<input type="checkbox"/>	<input type="checkbox"/>	Central America, South America	<input type="checkbox"/>	<input type="checkbox"/>	Cuba, Puerto Rico, other Caribbean	<input type="checkbox"/>	<input type="checkbox"/>	Asian (please mark specific geographic area below if known)	<input type="checkbox"/>	<input type="checkbox"/>	Japanese	<input type="checkbox"/>	<input type="checkbox"/>	Chinese	<input type="checkbox"/>	<input type="checkbox"/>	Korean	<input type="checkbox"/>	<input type="checkbox"/>	Vietnamese	<input type="checkbox"/>	<input type="checkbox"/>	Filipino	<input type="checkbox"/>	<input type="checkbox"/>	Native American	<input type="checkbox"/>	<input type="checkbox"/>	Other _____	<input type="checkbox"/>	<input type="checkbox"/>	Don't know
Father	Mother																																																																							
<input type="checkbox"/>	<input type="checkbox"/>	Caucasian (please mark specific geographic area below if known)																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Northern Europe (Austria, Denmark, Finland, France, Germany, Netherlands, Norway, Sweden, Switzerland, UK)																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Southern Europe (Greece, Italy, Spain, Turkey)																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Eastern Europe (Czechoslovakia, Hungary, Poland, Russia, Yugoslavia)																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Middle Eastern (Israel, Egypt, Iran, Iraq, Jordan, Syria, Other Arab States)																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	African-American																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Hispanic (please mark specific geographic area below if known)																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Mexico																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Central America, South America																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Cuba, Puerto Rico, other Caribbean																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Asian (please mark specific geographic area below if known)																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Japanese																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Chinese																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Korean																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Vietnamese																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Filipino																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Native American																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Other _____																																																																						
<input type="checkbox"/>	<input type="checkbox"/>	Don't know																																																																						
<table border="0" style="width: 100%;"> <tr> <td style="vertical-align: top; width: 15%;">How long have you lived there?</td> <td style="vertical-align: top; width: 15%;">Years</td> <td style="vertical-align: top; width: 30%;">What is your highest grade you completed in school?  <input type="checkbox"/> less than 8th grade  <input type="checkbox"/> 8th, 9th, 10th, or 11th grade  <input type="checkbox"/> high school graduate or equivalency  <input type="checkbox"/> some college, 2yr degree  <input type="checkbox"/> college graduate, 4yr degree  <input type="checkbox"/> post graduate education or degree               </td> <td style="vertical-align: top; width: 15%;">Mother Deceased? Cause of Death</td> <td style="vertical-align: top; width: 15%;">Mother: Father Deceased? Cause of Death Father:</td> </tr> <tr> <td></td> <td></td> <td></td> <td> <input type="checkbox"/> Yes  <input type="checkbox"/> No                  If Yes at what age?               </td> <td> <input type="checkbox"/> Yes  <input type="checkbox"/> No                  If Yes at what age?               </td> </tr> <tr> <td></td> <td></td> <td></td> <td>                 &lt; 29                  30-39                  40-49                  50-59                  60-69                  70-79                  80-89                  ≥ 90               </td> <td>                 &lt; 29                  30-39                  40-49                  50-59                  60-69                  70-79                  80-89                  &gt; 90               </td> </tr> <tr> <td></td> <td></td> <td></td> <td> <input type="checkbox"/> Heart Disease  <input type="checkbox"/> Cancer  <input type="checkbox"/> Stroke  <input type="checkbox"/> Accident  <input type="checkbox"/> Suicide  <input type="checkbox"/> Other, _____               </td> <td> <input type="checkbox"/> Heart Disease  <input type="checkbox"/> Cancer  <input type="checkbox"/> Stroke  <input type="checkbox"/> Accident  <input type="checkbox"/> Suicide  <input type="checkbox"/> Other, _____               </td> </tr> </table>													How long have you lived there?	Years	What is your highest grade you completed in school? <input type="checkbox"/> less than 8th grade <input type="checkbox"/> 8th, 9th, 10th, or 11th grade <input type="checkbox"/> high school graduate or equivalency <input type="checkbox"/> some college, 2yr degree <input type="checkbox"/> college graduate, 4yr degree <input type="checkbox"/> post graduate education or degree	Mother Deceased? Cause of Death	Mother: Father Deceased? Cause of Death Father:				<input type="checkbox"/> Yes <input type="checkbox"/> No If Yes at what age?	<input type="checkbox"/> Yes <input type="checkbox"/> No If Yes at what age?				< 29 30-39 40-49 50-59 60-69 70-79 80-89 ≥ 90	< 29 30-39 40-49 50-59 60-69 70-79 80-89 > 90				<input type="checkbox"/> Heart Disease <input type="checkbox"/> Cancer <input type="checkbox"/> Stroke <input type="checkbox"/> Accident <input type="checkbox"/> Suicide <input type="checkbox"/> Other, _____	<input type="checkbox"/> Heart Disease <input type="checkbox"/> Cancer <input type="checkbox"/> Stroke <input type="checkbox"/> Accident <input type="checkbox"/> Suicide <input type="checkbox"/> Other, _____																																								
How long have you lived there?	Years	What is your highest grade you completed in school? <input type="checkbox"/> less than 8th grade <input type="checkbox"/> 8th, 9th, 10th, or 11th grade <input type="checkbox"/> high school graduate or equivalency <input type="checkbox"/> some college, 2yr degree <input type="checkbox"/> college graduate, 4yr degree <input type="checkbox"/> post graduate education or degree	Mother Deceased? Cause of Death	Mother: Father Deceased? Cause of Death Father:																																																																				
			<input type="checkbox"/> Yes <input type="checkbox"/> No If Yes at what age?	<input type="checkbox"/> Yes <input type="checkbox"/> No If Yes at what age?																																																																				
			< 29 30-39 40-49 50-59 60-69 70-79 80-89 ≥ 90	< 29 30-39 40-49 50-59 60-69 70-79 80-89 > 90																																																																				
			<input type="checkbox"/> Heart Disease <input type="checkbox"/> Cancer <input type="checkbox"/> Stroke <input type="checkbox"/> Accident <input type="checkbox"/> Suicide <input type="checkbox"/> Other, _____	<input type="checkbox"/> Heart Disease <input type="checkbox"/> Cancer <input type="checkbox"/> Stroke <input type="checkbox"/> Accident <input type="checkbox"/> Suicide <input type="checkbox"/> Other, _____																																																																				

FIG. 22A



[illegible]

**FIG. 22C**

**SUBSTITUTE SHEET (RULE 26)**

What is your highest grade  
you completed in school?

- ☐ less than 8th grade  
☐ 8th,9th,10th,or11th grade  
☐ high school graduate or equivalency  
☐ some college, 2yr degree  
☐ college graduate,4yr degree  
☐ post graduate education or degree

**Mother Deceased? Cause of Death Mother: Father Deceased? Cause of Death Father:**

- ☐ Yes  
☐ No

**If Yes at  
what age?**

- $\leq 29$   
 30-39  
 40-49  
 50-59  
 60-69  
 70-79  
 80-89  
 $\geq 90$

- ☐ Heart Disease  
☐ Cancer  
☐ Stroke  
☐ Accident  
☐ Suicide  
☐ Other, \_\_\_\_\_

☐ Yes  
☐ No

If Yes at  
what age?

- $\leq 29$   
 30-39  
 40-49  
 50-59  
 60-69  
 70-79  
 80-89  
 $> 90$

- ☐ Heart Disease  
☐ Cancer  
☐ Stroke  
☐ Accident  
☐ Suicide  
☐ Other \_\_\_\_\_

## Health Information

Have you or has anyone in your immediate family (parents, brothers, sisters, or your children) had the following?

**Mark all that apply!**

## Disease

	10	mother	father	sister	brother	child
Heart Disease	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stroke	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hypertension	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Blood clots	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diabetes, insulin dependent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diabetes, not insulin-dependent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cancer:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lung&Bronchus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Breasts	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Prostate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Colon&Rectum	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Skin	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lymphoma&Leukemia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other, please specify below:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Alzheimer's Disease	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Epilepsy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Schizophrenia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bipolar disorder (manic depression)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Major depression	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Chronic Inflammatory or Autoimmune Disease including Multiple Sclerosis and Rheumatoid Arthritis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Emphysema	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Asthma	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other, please specify below:						

**Do you take prescription drugs on a regular basis?**

☐ Yes      ☐ No

Have you ever donated ☐ Yes ☐ No  
blood before?

If yes, please specify below:

If yes, how many times:

Number of Times

Have you been hospitalized  
in the past 5 years for more  
than 6 days at a time?

☐ Yes      ☐ No

1) Weeks: ☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6

- ☐ Acute disorder, including infection and thrombosis  
☐ Chronic Disorder  
☐ Accident  
☐ Other: \_\_\_\_\_

2) Weeks: ☐ ☒ ☐ ☐ ☐ ☐ ☐ ☐

- ☐ Acute disorder, including infection and thrombosis  
☐ Chronic Disorder  
☐ Accident  
☐ Other: \_\_\_\_\_

3) Weeks:

- ☐ Acute disorder, including infection and thrombosis  
☐ Chronic Disorder  
☐ Accident  
☐ Other: \_\_\_\_\_

**Do you drink any kind of alcoholic beverage?**

- ☐ Never ☐ Hardly ever  
☐ Less than 3 times per week ☐ 3 or more times per week  
☐ Daily

**Additional health information details you would like to provide:**

FOR OFFICE  
USE ONLY

FIG. 22D

**SUBSTITUTE SHEET (RULE 26)**

32 / 51

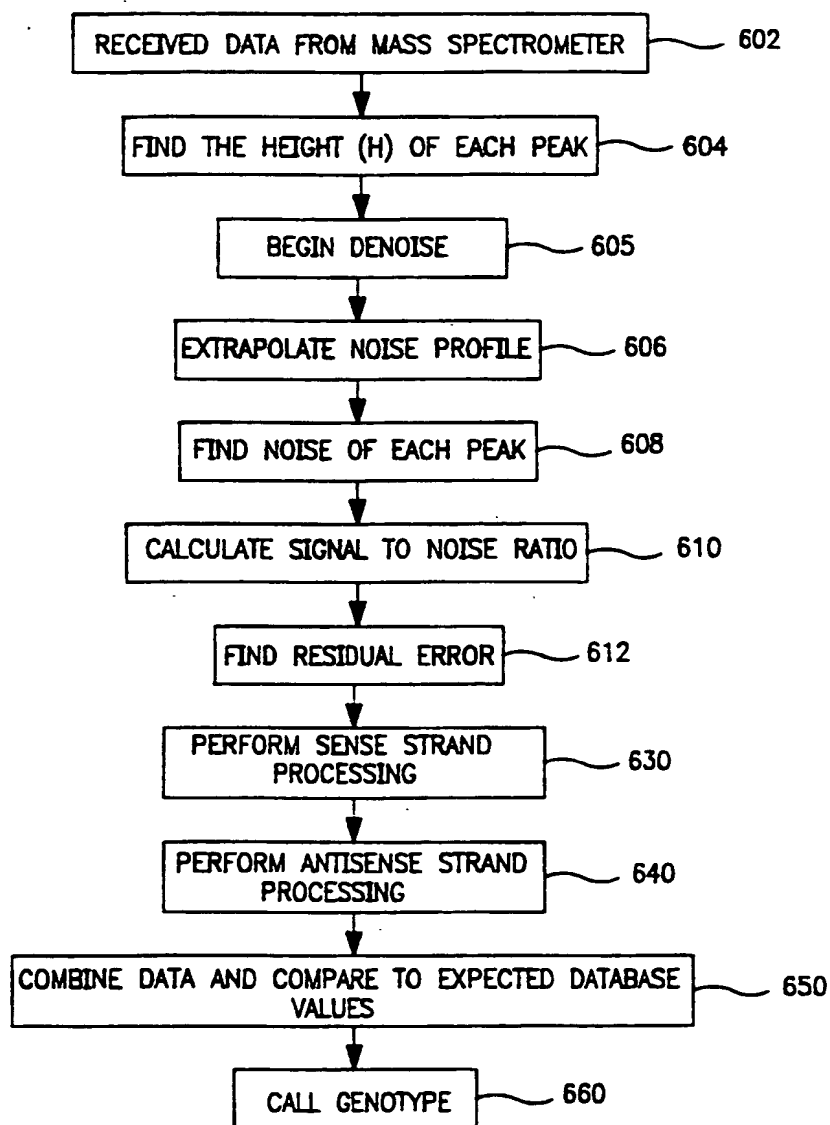


FIG. 23

33/ 51

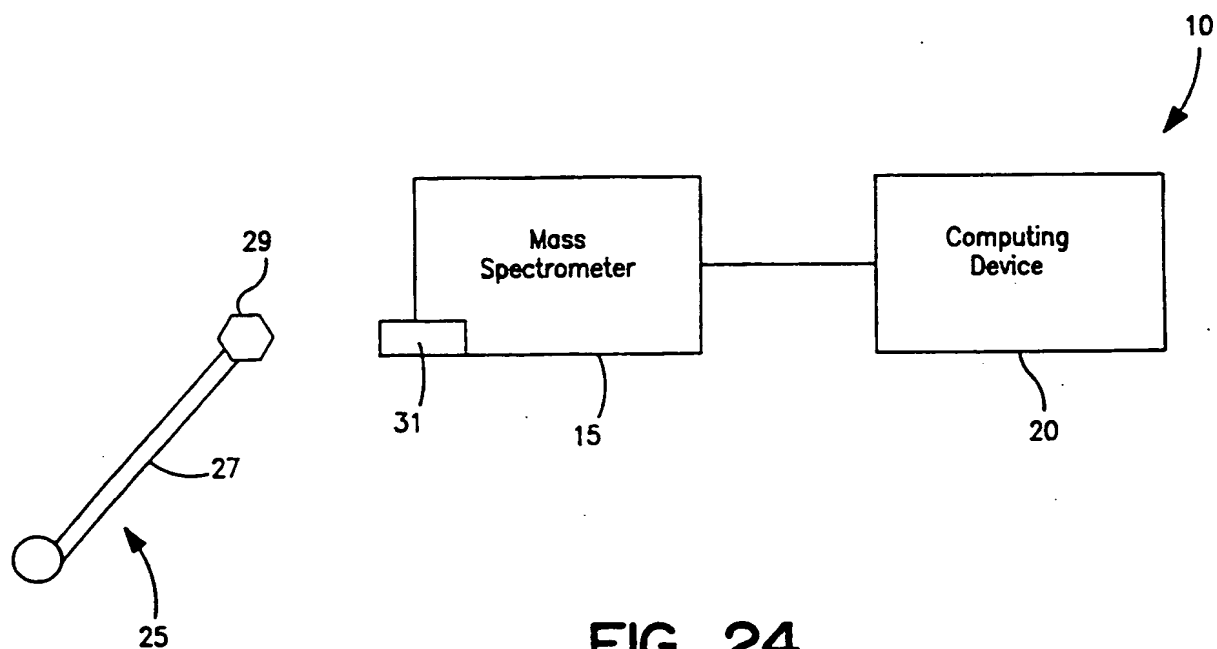
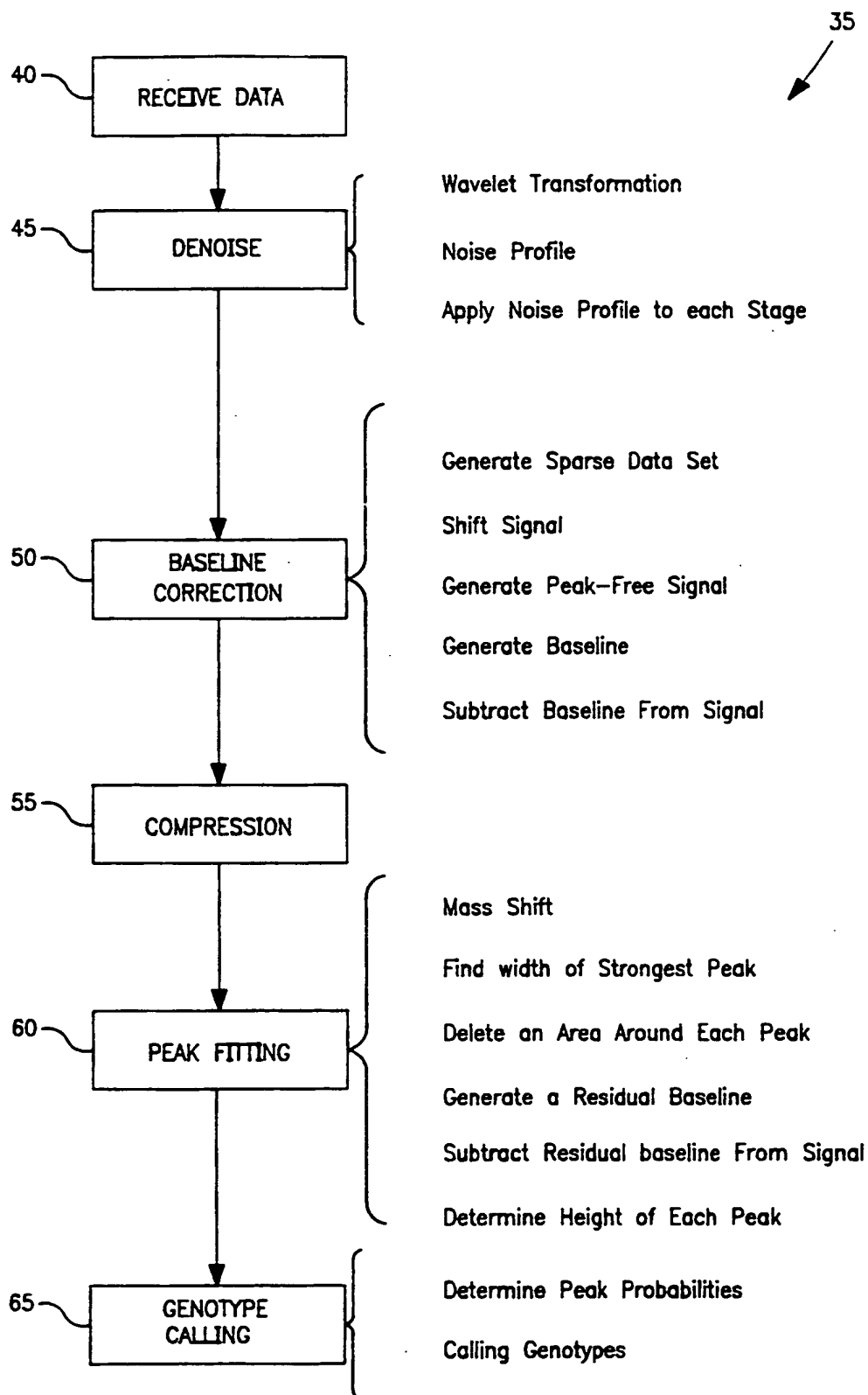


FIG. 24

34 / 51

**FIG. 25**

SUBSTITUTE SHEET (RULE 26)

35/51

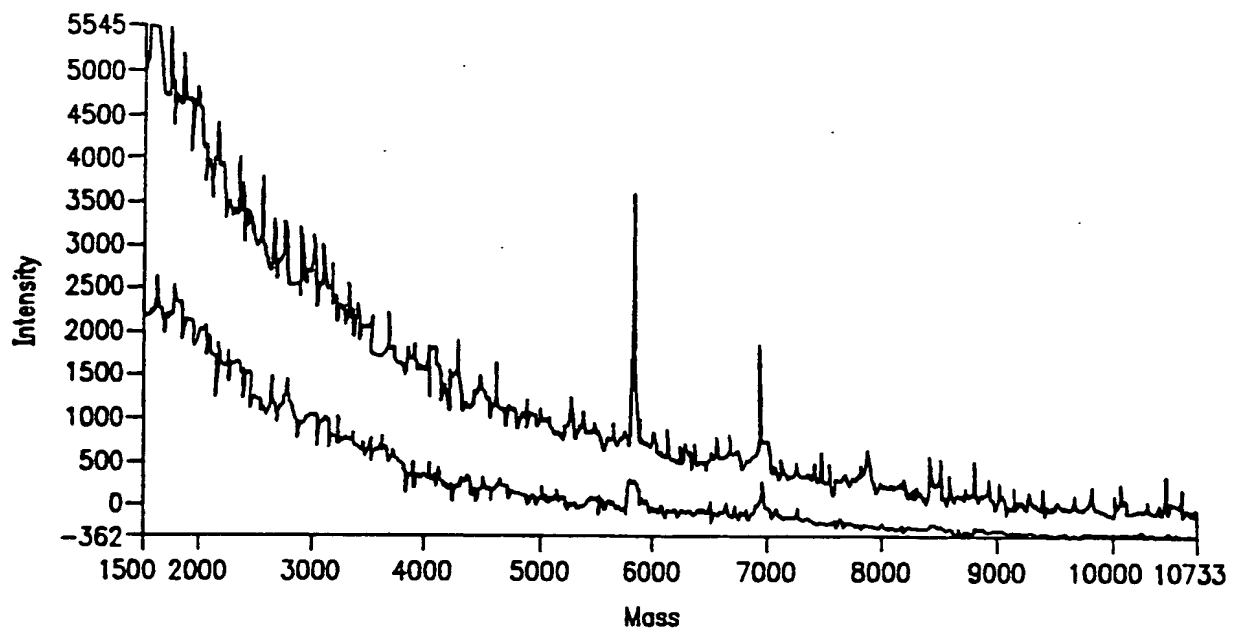


FIG. 26

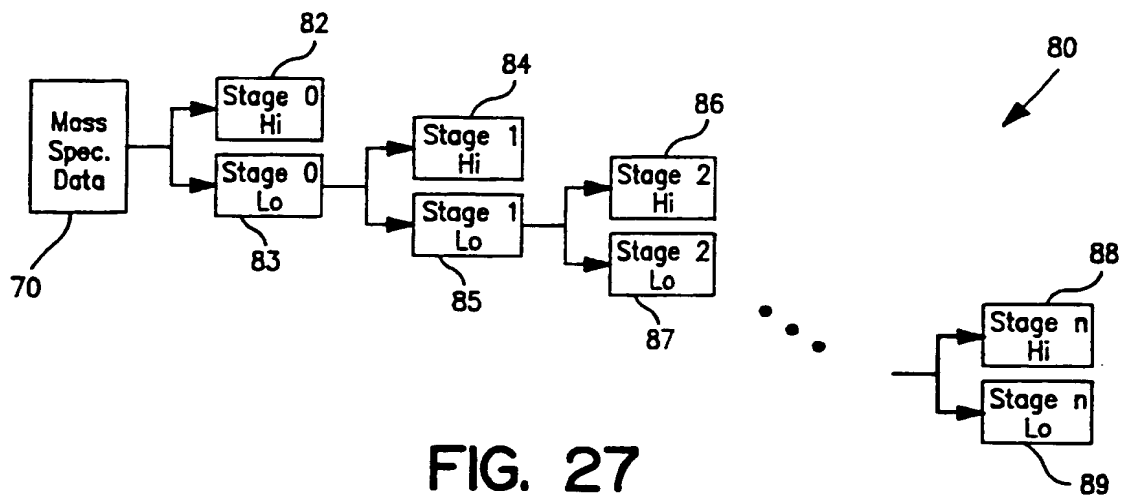


FIG. 27

36/51

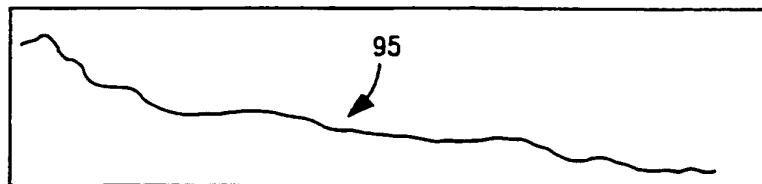


FIG. 28

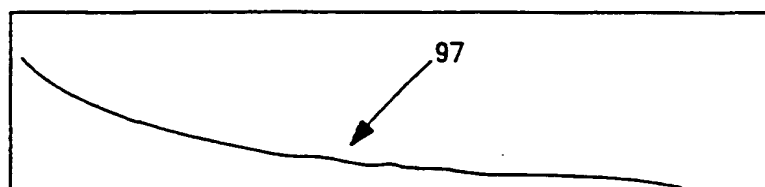


FIG. 29

Exp fitting  
 $a_0 + a_1 \exp^{-a_2 m}$

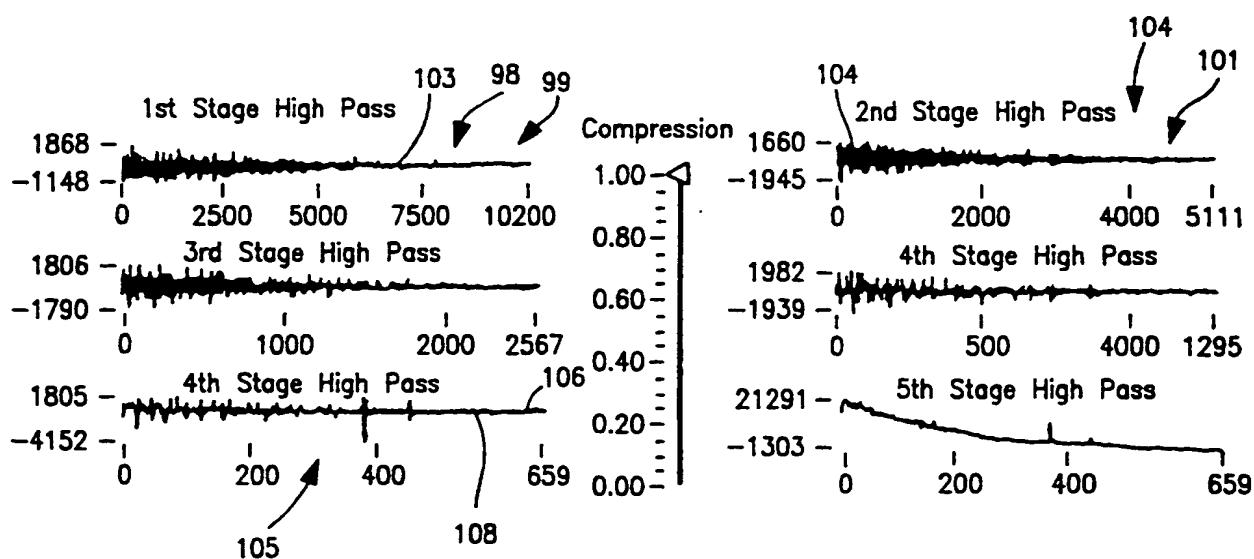
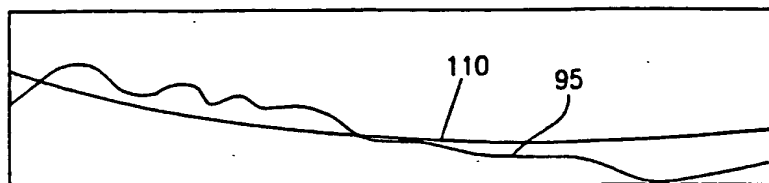


FIG. 30

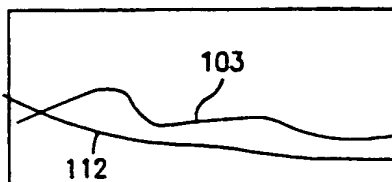
37/ 51

Stage 0 - Hi



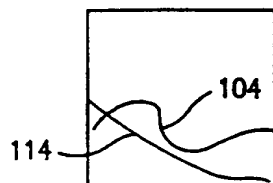
Threshold 0=4XNoiseProfile

Stage 1 - Hi



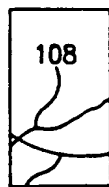
Threshold 1=2XNoiseProfile

Stage 2 - Hi



Threshold 2=1XNoiseProfile

Stage n - Hi



Threshold  $n=(1/2^{n-2})XNoiseProfile$

116

Stage n - Lo

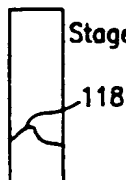


FIG. 31

38/51

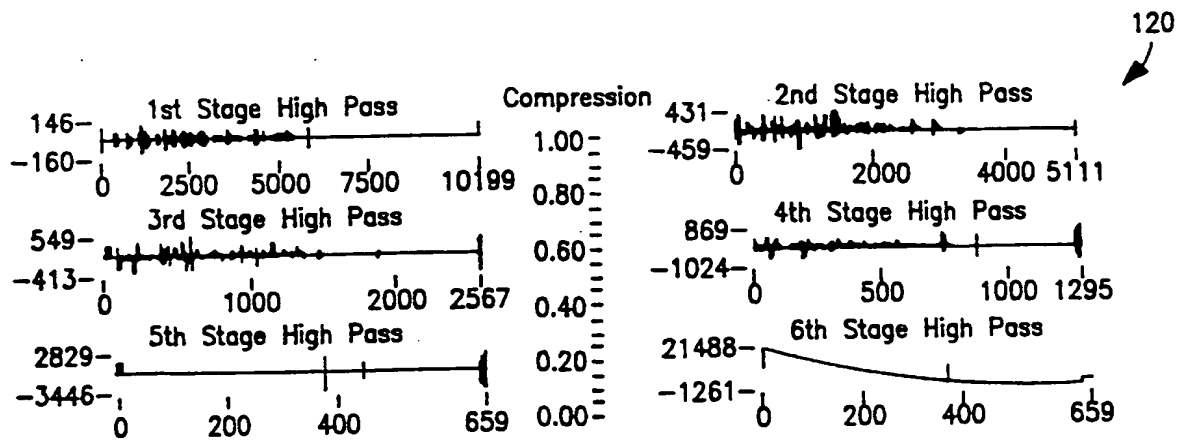


FIG. 32

$$\text{Signal } (t) = \frac{\text{Start } 0(t) + \text{Start } 1(t) + \text{Start } 2(t) \dots + \text{Start } 23(t)}{24}$$

SHIFT SIGNAL TO ACCOUNT FOR  
VARIATIONS DUE TO STARTING POINT

FIG. 33

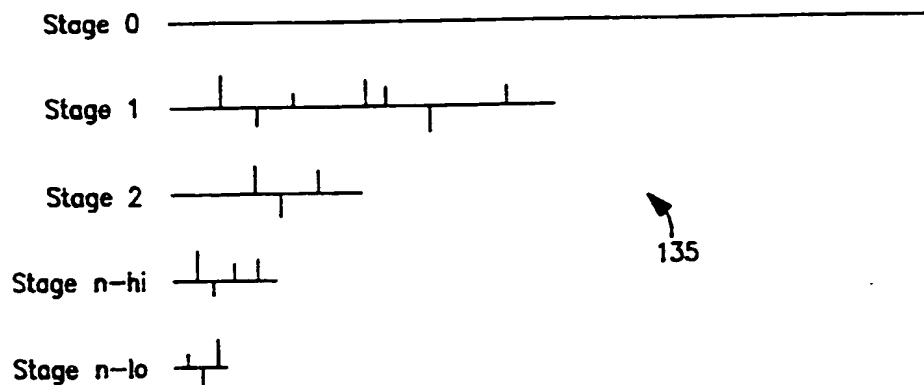


FIG. 34

SUBSTITUTE SHEET (RULE 26)

39/51

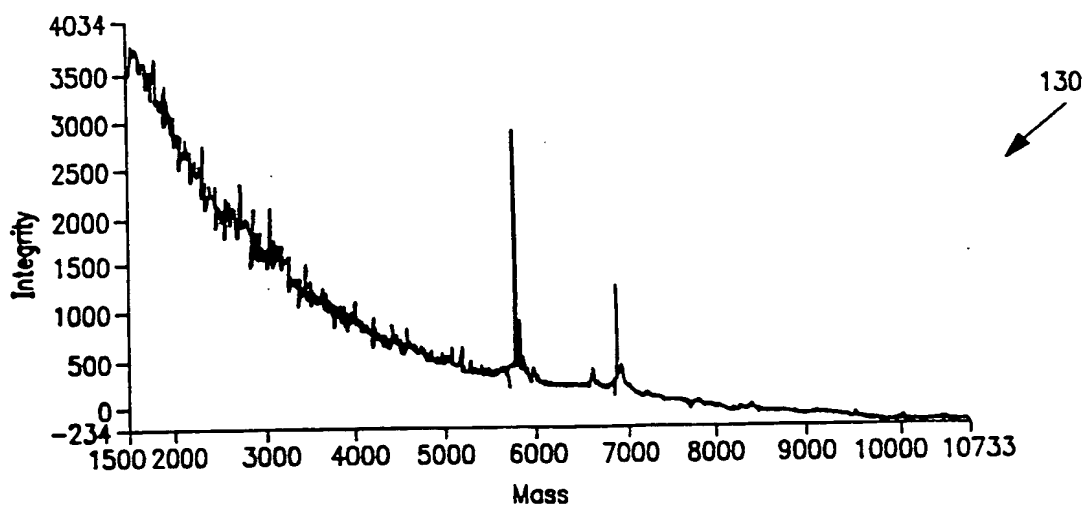


FIG. 35

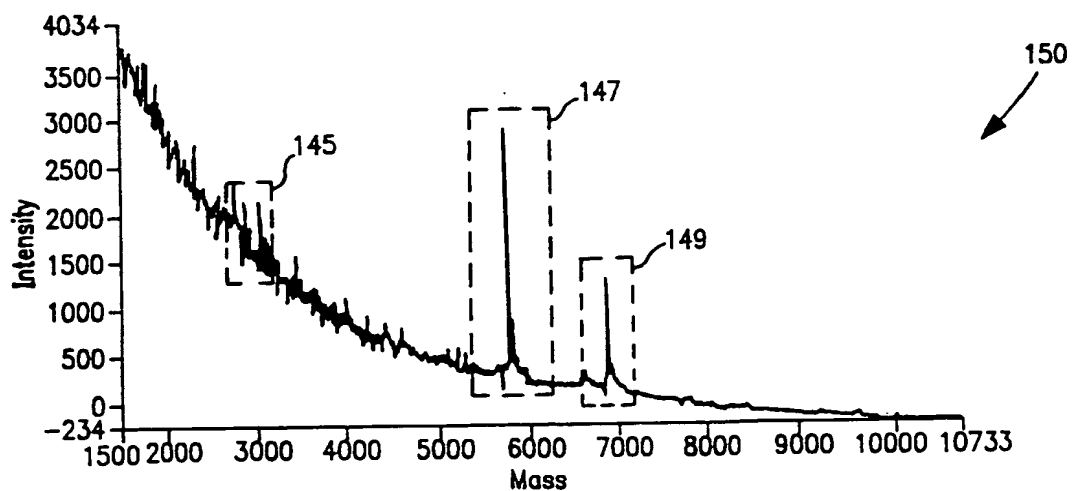


FIG. 13—TAKE A MOVING AVERAGE, REMOVE SECTIONS EXCEEDING A THRESHOLD

FIG. 36

40/51

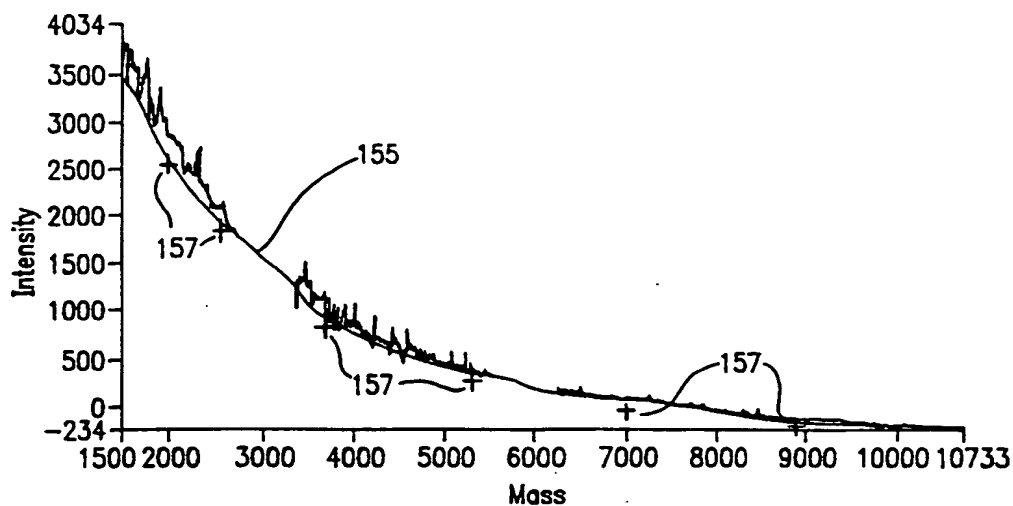


FIG. 37

FIND MINIMA IN REMAINING SIGNALS AND CONNECT TO  
FORM A PEAK FREE SIGNAL

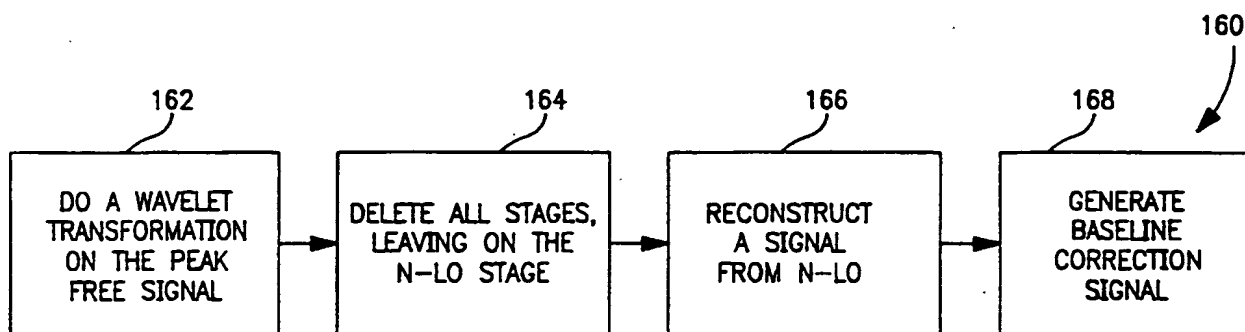


FIG. 38

GENRATE BASELINE CORRECTION

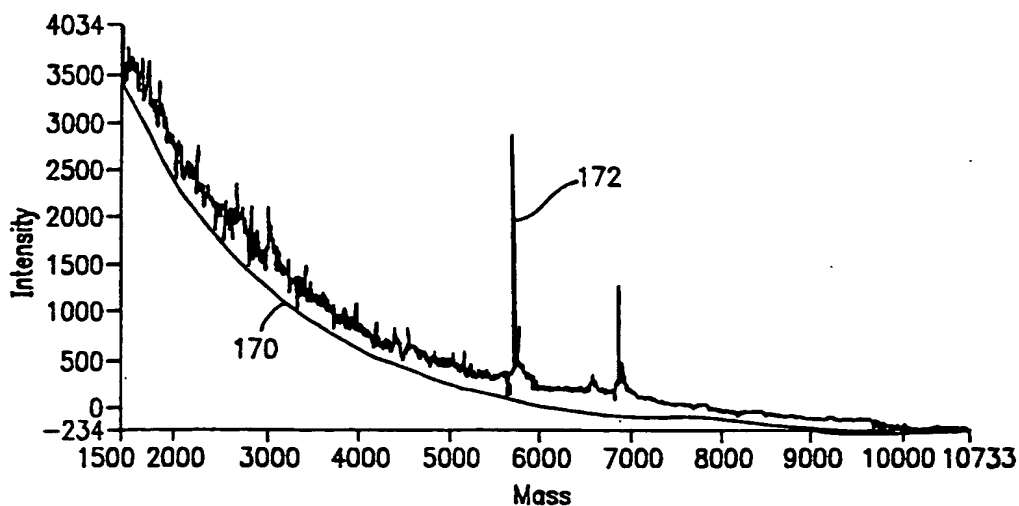


FIG. 39

SUBSTITUTE SHEET (RULE 26)

41/51

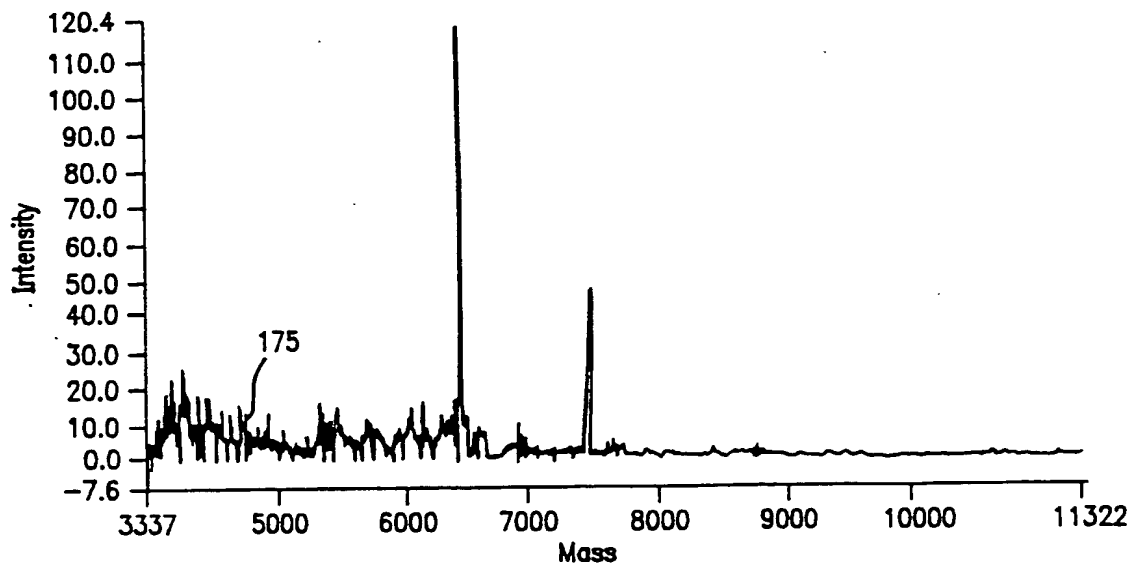


FIG. 40

NON-O COEFFICIENTS	COEFFICIENTS	INTERMEDIATE	RELATIVE
183 → 100 150 500 10,050 10,075 11,125 12,100 13,250	182 → 25 220 .1 800 890 910 1000 (MAX) 940	186 → 188 → 190 → 100.025 150.220 500.0001 10,050.8 10,075.89 11,125.91 12,100.99999 13,250.94	180 → 195 → 100.025 50.220 350.0001 9550.8 25.89 150.91 975.99999 1150.94

FIG. 41

42 / 51

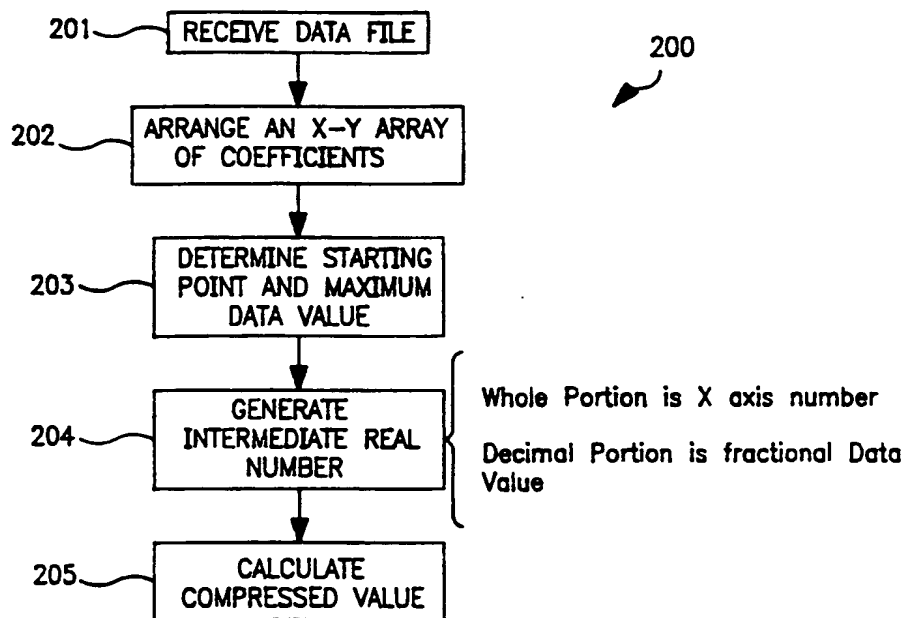


FIG. 42

43/51

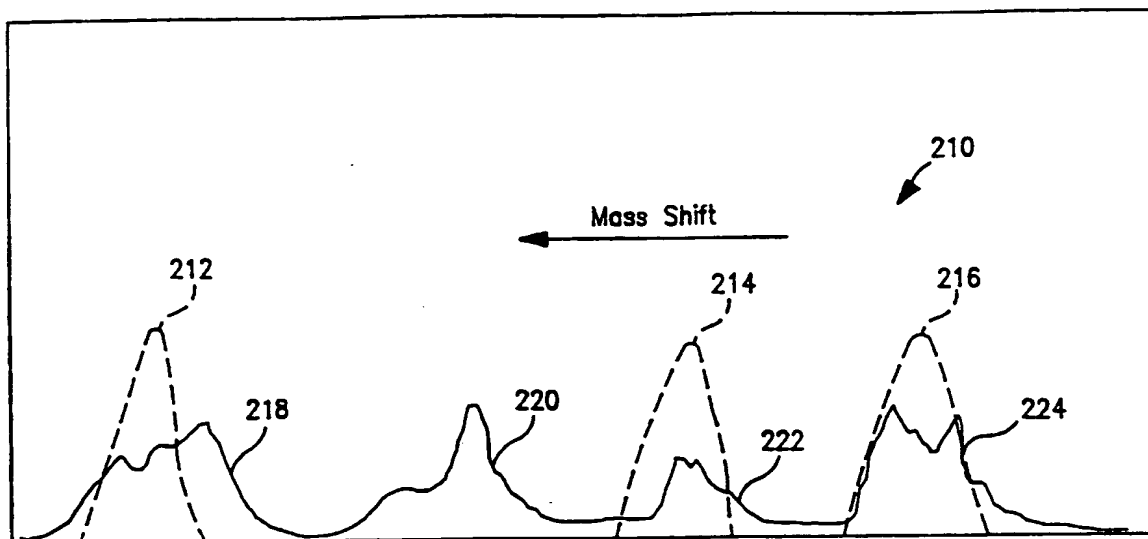


FIG. 43

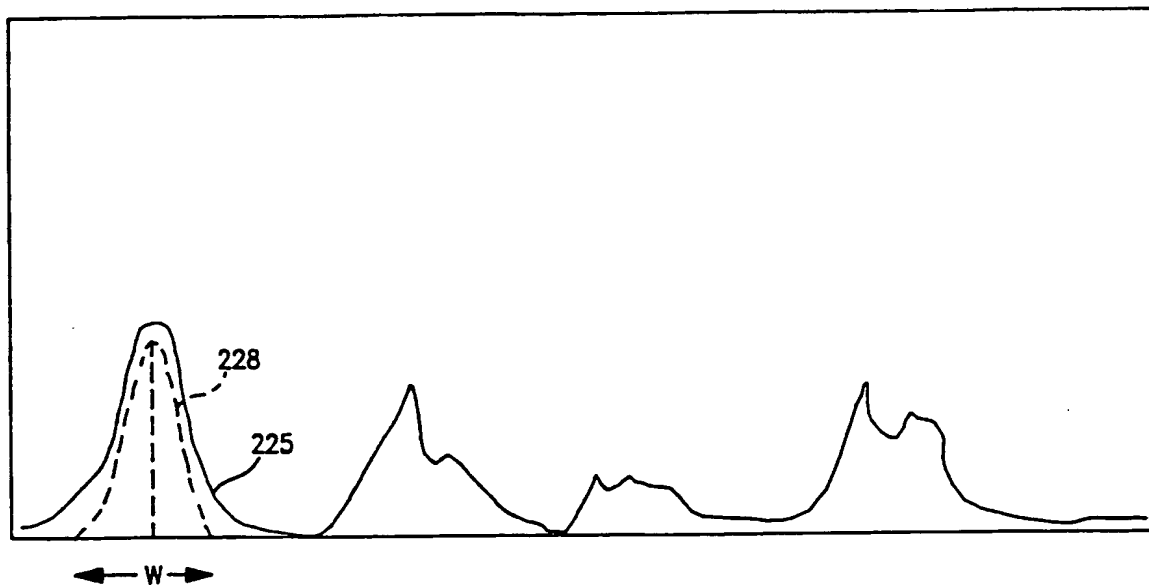


FIG. 44

44/51

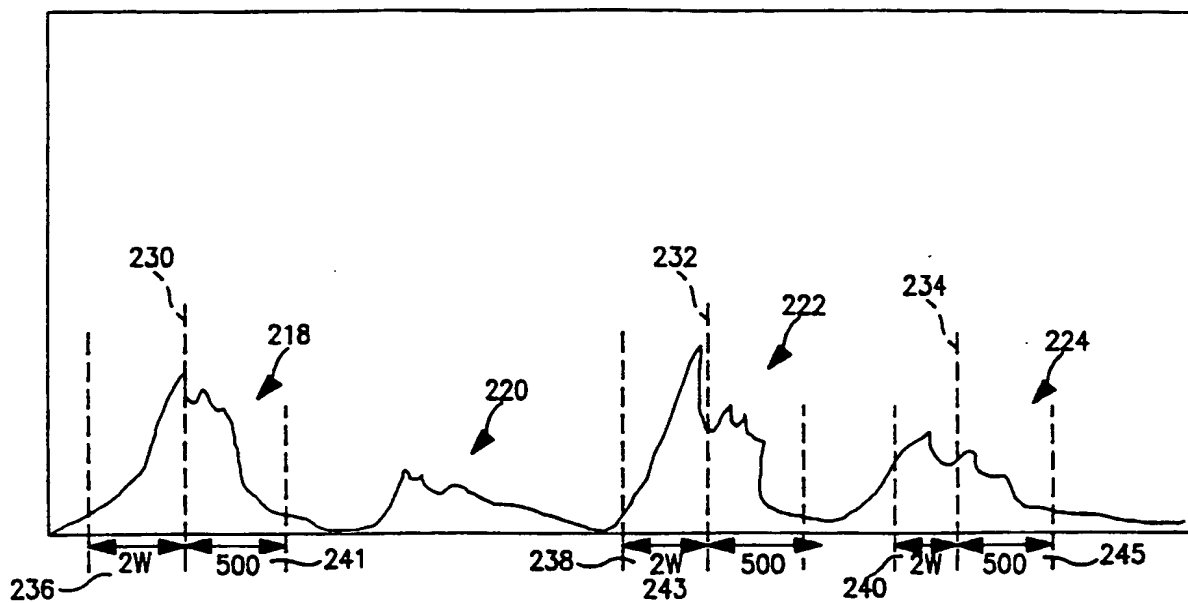


FIG. 45

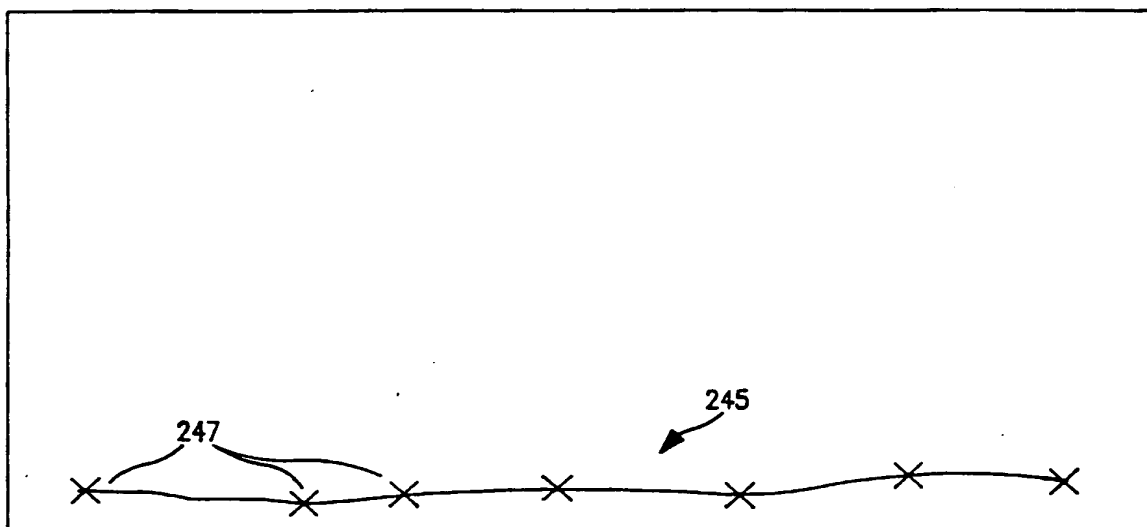


FIG. 46

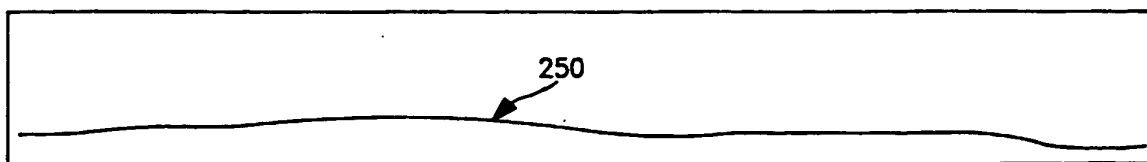


FIG. 47

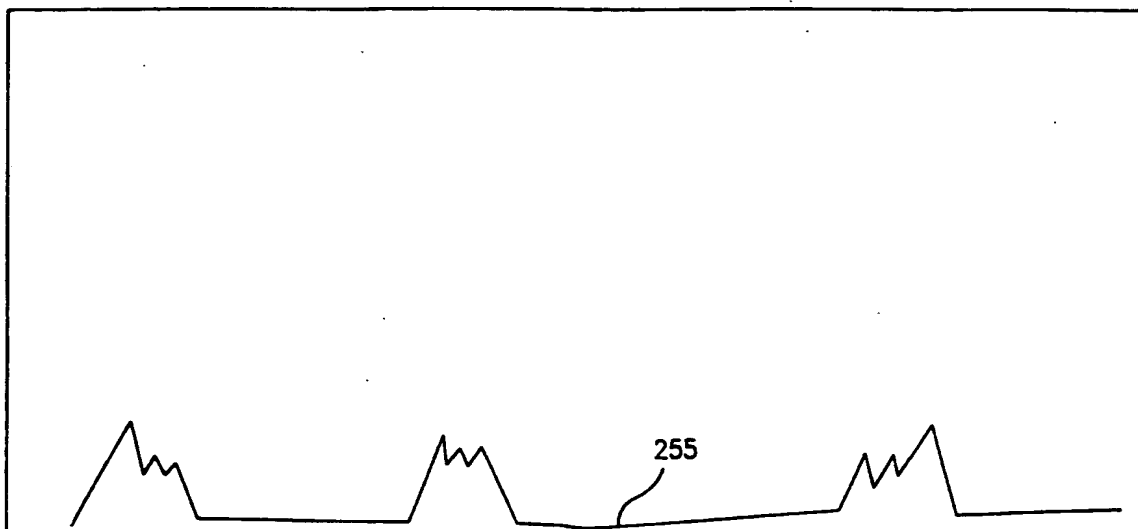


FIG. 48

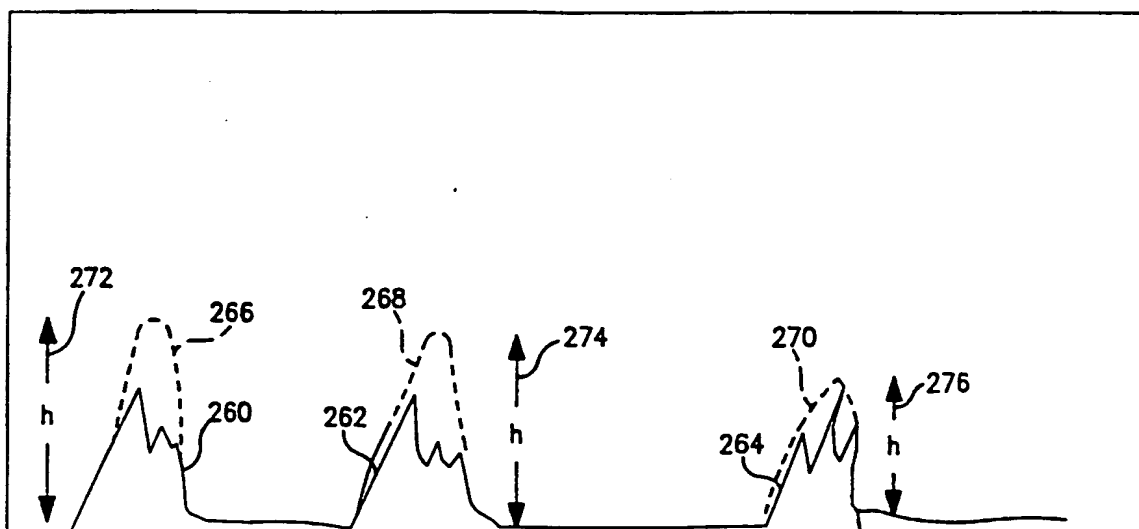


FIG. 49

46/51

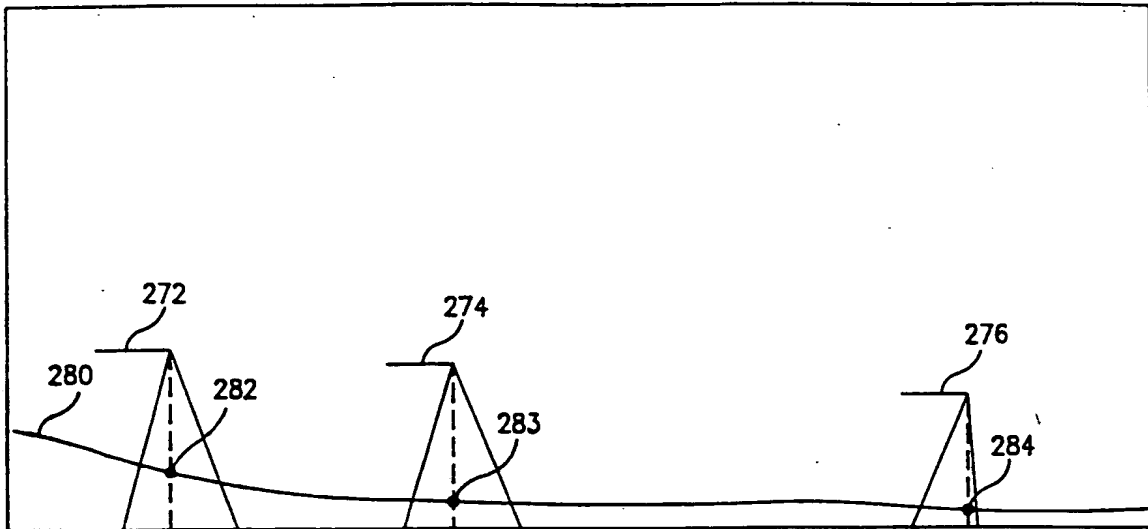


FIG. 50

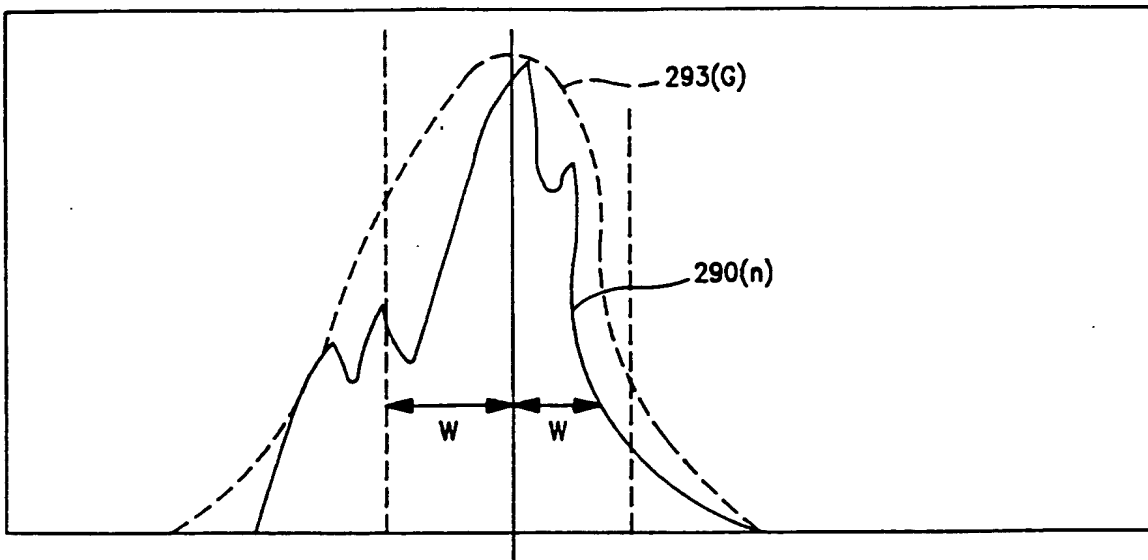
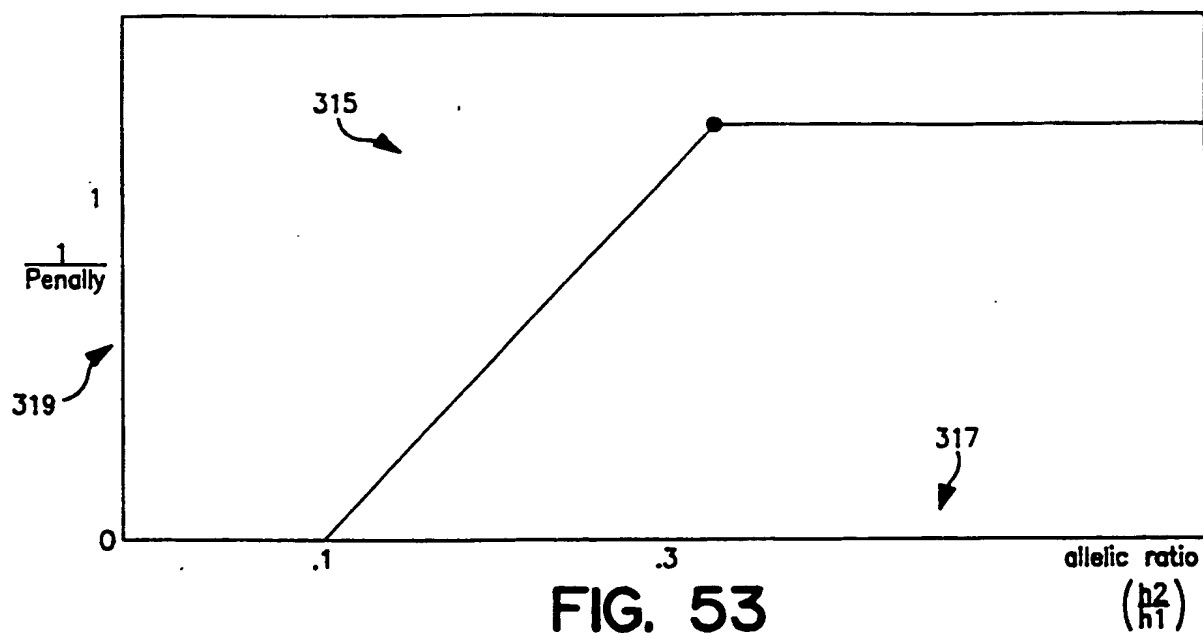
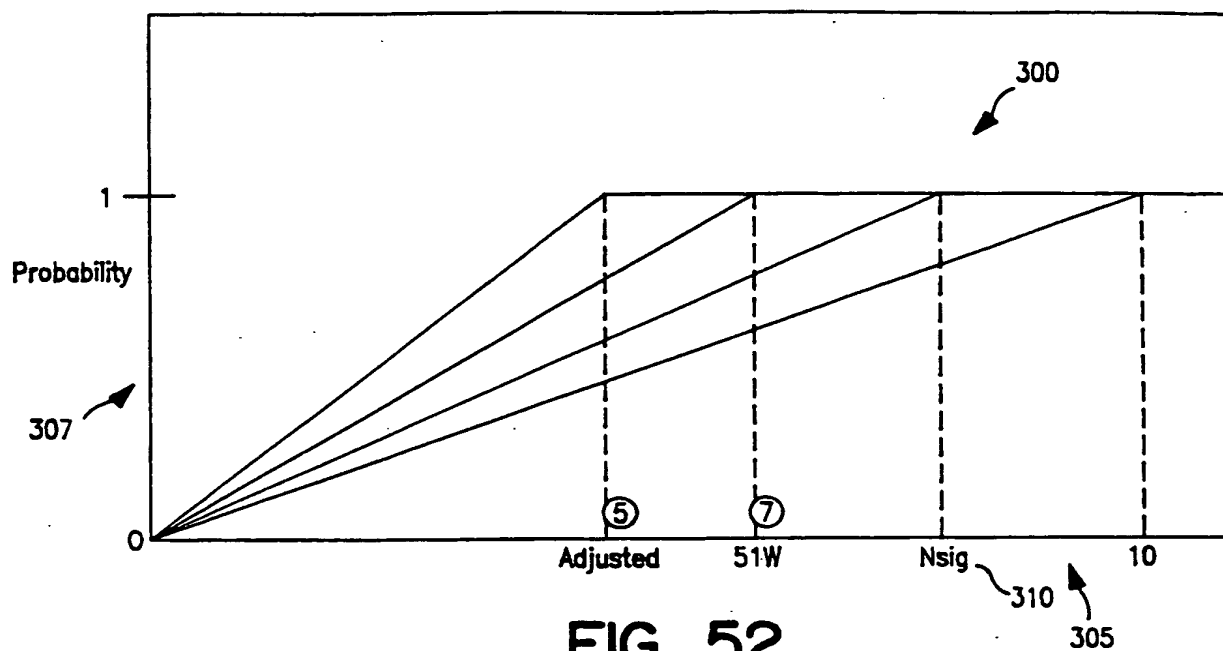


FIG. 51

47/51



48/51

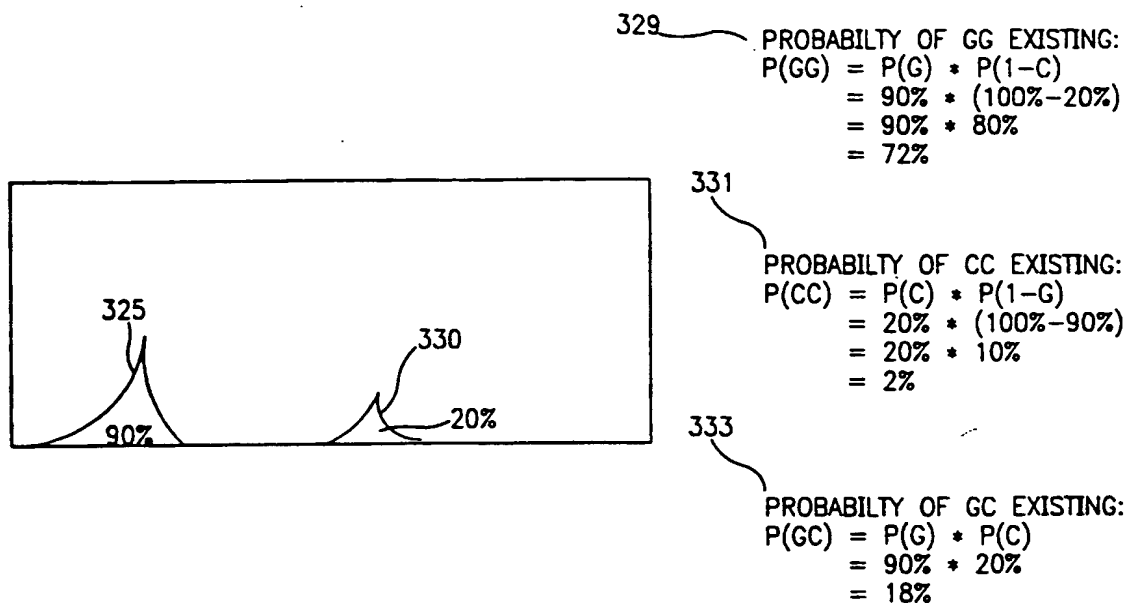


FIG. 54

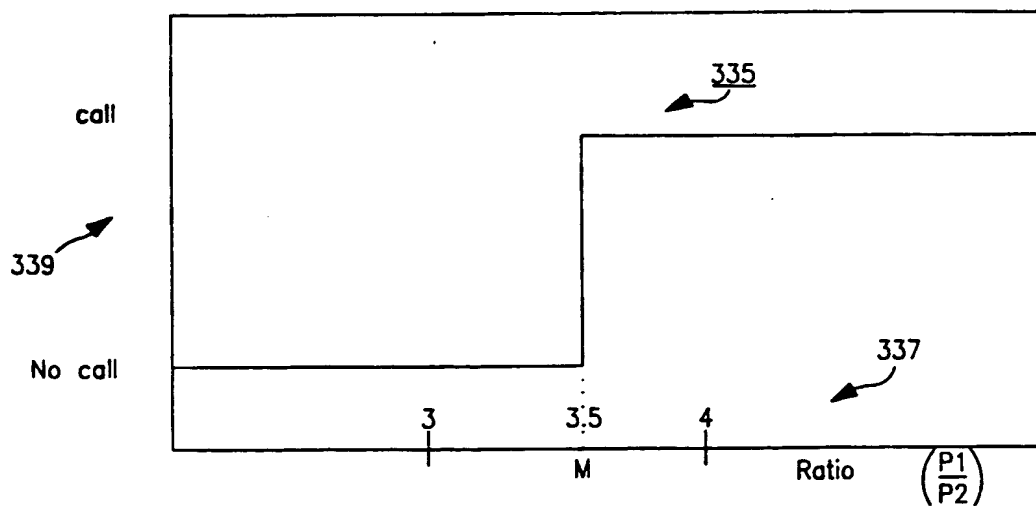


FIG. 55

49/51

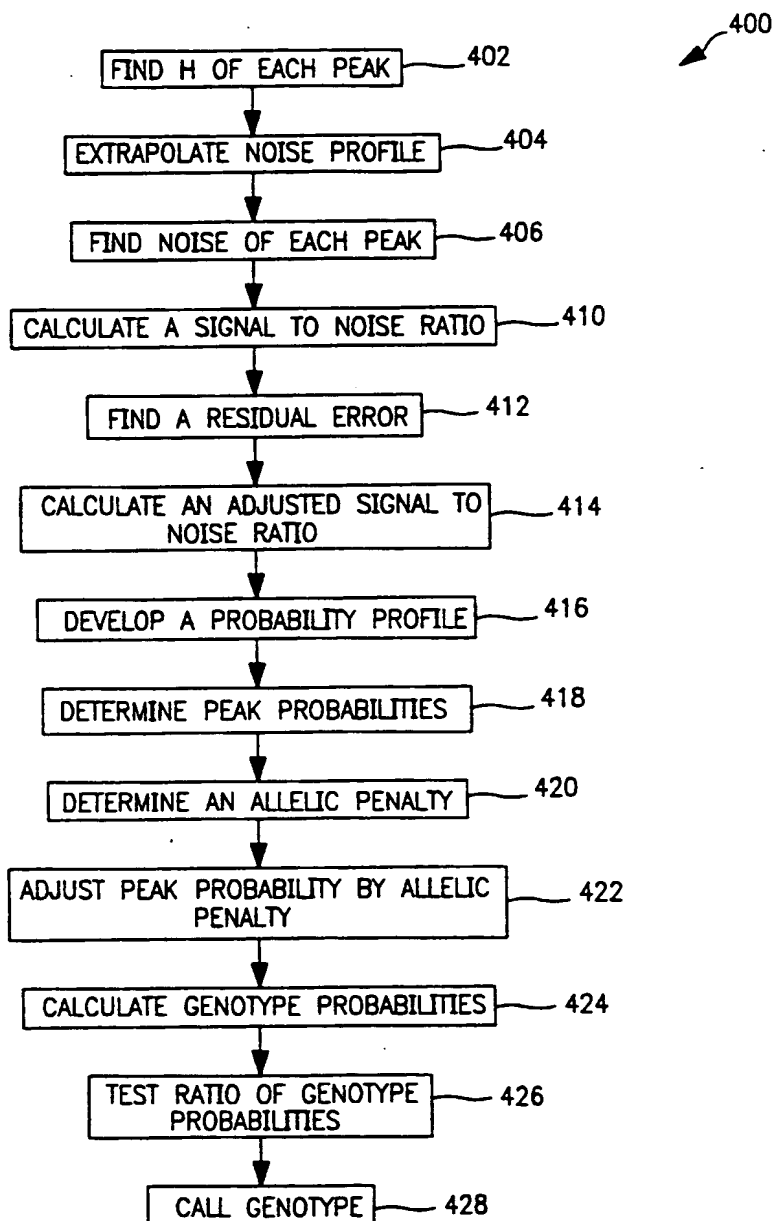


FIG. 56

50/ 51

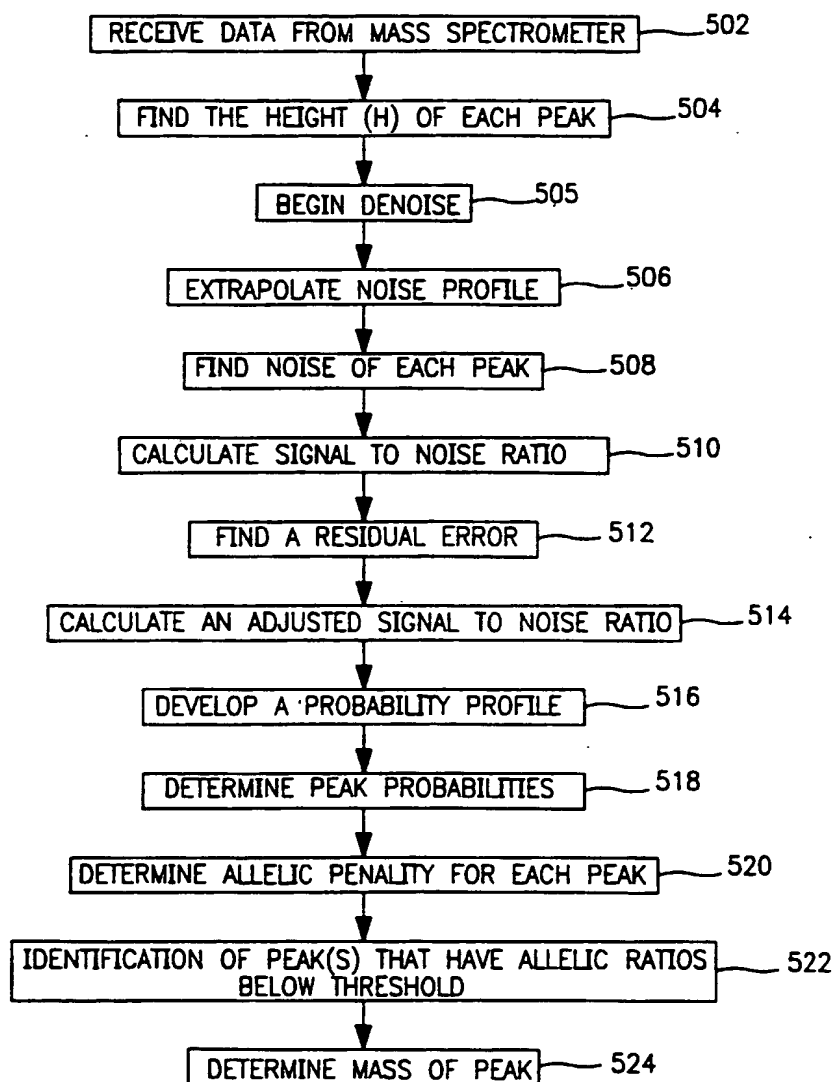


FIG. 57

51/51

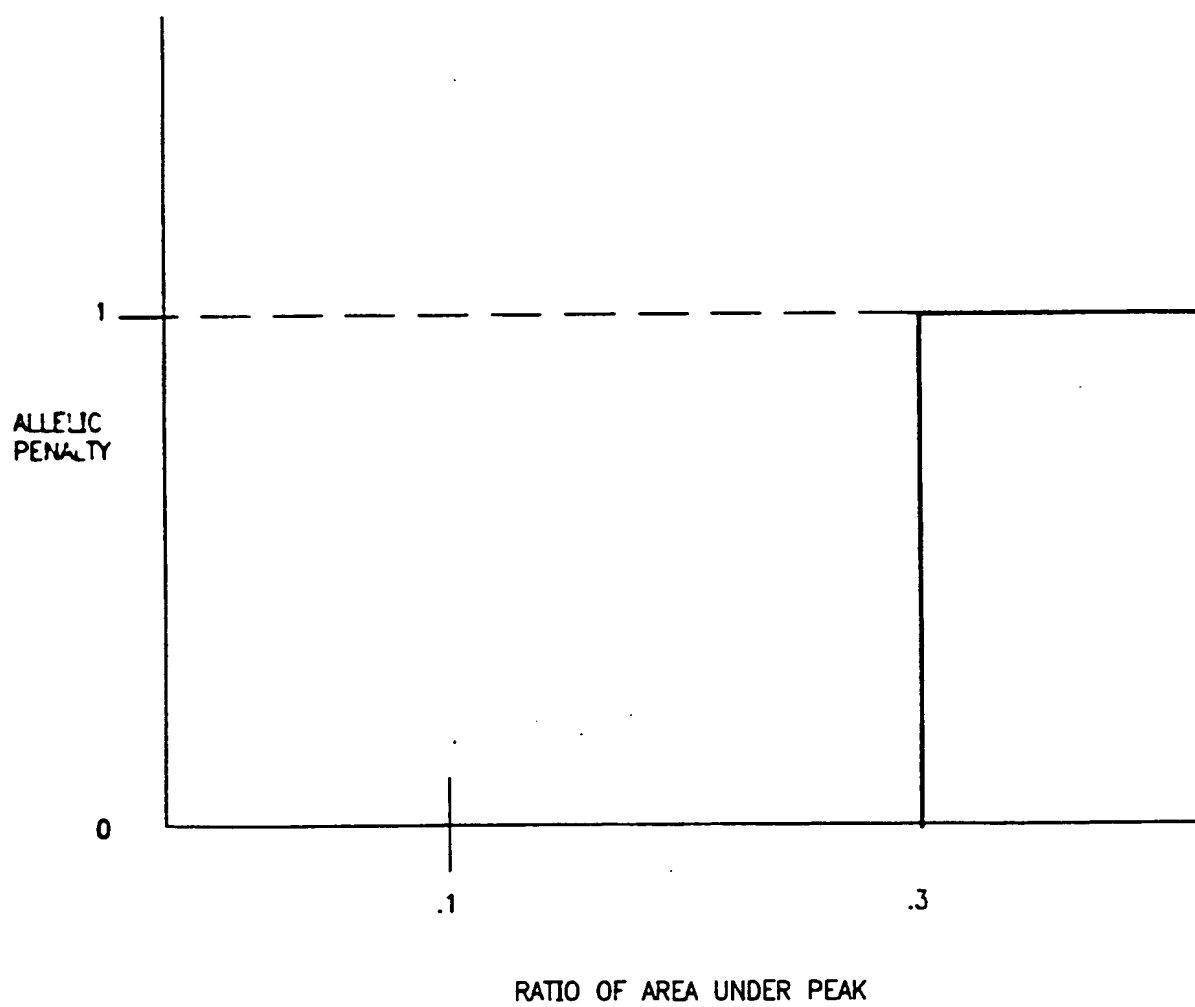


FIG. 58

## SEQUENCE LISTING

<110> SEQUENOM  
Braun et al.

<120> METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING  
POLYMORPHIC GENETIC MARKERS

<130> 24736-2033PC

<140> Not Yet Assigned

<141> 2000-10-13

<150> 60/217,658

<151> 2000-07-10

<150> 60/159,176

<151> 1999-10-13

<150> 60/217,251

<151> 2000-07-10

<150> 09/663,968

<151> 2000-09-19

<160> 118

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 361

<212> DNA

<213> Homo Sapien

<400> 1

ctgaggacct	ggtcctctga	ctgctctttt	cacccatcta	cagtcctccct	tgccgtccca	60
agcaatggat	gatttgatgc	tgtccccgga	cgatattgaa	caatgggttca	ctgaagaccc	120
aggtccagat	gaagctccca	gaatgccaga	ggctgctccc	cgcgtggccc	ctgcaccagc	180
agctcctaca	ccggcggccc	ctgcaccagc	cccctcctgg	cccctgtcat	cttctgtccc	240
ttcccagaaa	acctaccagg	gcagctacgg	tttccgtctg	ggcttcttgc	attctgggac	300
agccaagtct	gtgacttgca	cggtcagttg	ccctgagggg	ctggcttcca	tgagacttca	360
a						361

<210> 2

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide Primer

<400> 2

cccagtcacg	acgttgtaaa	acgctgagga	cctggctcctc	tgac	44
------------	------------	------------	-------------	------	----

<210> 3

<211> 42

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide Primer

<400> 3

agcggataac	aatttcacac	aggttgaagt	ctcatggaag	cc	42
------------	------------	------------	------------	----	----

<210> 4

<211> 17

<212> DNA

2/122

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 4

gccagaggct gctcccc

17

&lt;210&gt; 5

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 5

gccagaggct gctcccc

17

&lt;210&gt; 6

&lt;211&gt; 19

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 6

gccagaggct gctccccgc

19

&lt;210&gt; 7

&lt;211&gt; 18

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 7

gccagaggct gctcccc

18

&lt;210&gt; 8

&lt;211&gt; 161

&lt;212&gt; DNA

&lt;213&gt; Homo Sapien

&lt;400&gt; 8

gtccgtcaga acccatgcgg cagcaaggcc tgccgccgcc tcttcggccc agtggacagc  
 gagcagctga gccgcgactg tgatgcgcta atggcgggct gcatccagga ggcccgtgag  
 cgatggaact tcgactttgt caccgagaca ccactggagg g

60  
 120  
 161

&lt;210&gt; 9

&lt;211&gt; 43

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligonucleotide Primer

&lt;400&gt; 9

3/122

cccagtcacg acgttgtaaa acggtccgtc agaaccatg cgg	43
<210> 10	
<211> 44	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Oligonucleotide Primer	
<400> 10	
agcggataac aatttcacac aggctccagt ggtgtctcgg tgac	44
<210> 11	
<211> 15	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Oligonucleotide Primer	
<400> 11	
cagcgagcag ctgag	15
<210> 12	
<211> 15	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Probe	
<400> 12	
cagcgagcag ctgag	15
<210> 13	
<211> 16	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Probe	
<400> 13	
cagcgagcag ctgagc	16
<210> 14	
<211> 17	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Probe	
<400> 14	
cagcgagcag ctgagac	17
<210> 15	
<211> 205	
<212> DNA	

4/122

&lt;213&gt; Homo Sapien

&lt;400&gt; 15

gcgctccatt catctcttca tcgactctct gttgaatgaa gaaaatccaa gtaaggccta	60
caggtgcagt tccaaggaag cctttgagaa agggctctgc ttgagttgta gaaagaaccg	120
ctgcaacaat ctgggctatg agatcaataa agtcagagcc aaaagaagca gcaaaatgta	180
cctgaagact cgttctcaga tgccc	205

&lt;210&gt; 16

&lt;211&gt; 42

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligonucleotide Primers

&lt;400&gt; 16

cccagtcacg acgttgtaaa acggcgctcc attcatctct tc	42
--	----

&lt;210&gt; 17

&lt;211&gt; 42

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligonucleotide Primer

&lt;400&gt; 17

agcggataac aatttcacac agggggcatc tgagaacgag tc	42
--	----

&lt;210&gt; 18

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligonucleotide Primer

&lt;400&gt; 18

caatctgggc tatgagatca	20
-----------------------	----

&lt;210&gt; 19

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 19

caatctgggc tatgagatca	20
-----------------------	----

&lt;210&gt; 20

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

5/122

<400> 20  
 caatctgggc tatgagatca a 21  
  
 <210> 21  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Probe  
  
 <400> 21  
 caatctgggc tatgagatca gt 20  
  
 <210> 22  
 <211> 60  
 <212> DNA  
 <213> Homo Sapien  
  
 <220>  
 <223> Probe  
  
 <400> 22  
 gtgccggcta ctcggatggc agcaaggact cctgcaaggg ggacagtgga ggcccacatg 60  
  
 <210> 23  
 <211> 60  
 <212> DNA  
 <213> Homo sapien  
  
 <400> 23  
 ccacccacta ccggggcacg tggtaacctga cgggcatcgt cagctggggc cagggctgcg 60  
  
 <210> 24  
 <211> 42  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Oligonucleotide primer  
  
 <400> 24  
 cccagtcacg acgttgtaaa acgatggcag caaggactcc tg 42  
  
 <210> 25  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Oligonucleotide primer  
  
 <400> 25  
 cacatgccac ccactacc 18  
  
 <210> 26  
 <211> 43  
 <212> DNA  
 <213> Artificial Sequence

6/122

<220>  
<223> Oligonucleotide primer

<400> 26  
agcggataac aatttcacac aggtgacgat gcccgtcagg tac 43

<210> 27  
<211> 15  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 27  
atgccaccca ctacc 15

<210> 28  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 28  
cacatgccac ccactaccg 19

<210> 29  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 29  
cacatgccac ccactaccag 20

<210> 30  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 30  
agcggataac aatttcacac agg 23

<210> 31  
<211> 2363  
<212> DNA  
<213> Homo Sapien

<220>  
<221> CDS  
<222> (138)...(2126)  
<223> AKAP-10

7/122

&lt;300&gt;

&lt;308&gt; GenBank AF037439

&lt;309&gt; 1997-12-21

&lt;400&gt; 31

```

gcggccttggt gataatatgg cggctggagc tgcctgggca tcccaggag gcggtggggc      60
ccactcccgg aagaagggtc ccttttcgcg ctagtgcagc ggccccctctg gacccggaag      120
tccggggccgg ttgctga atg agg gga gcc ggg ccc tcc ccg cgc cag tcc      170
                Met Arg Gly Ala Gly Pro Ser Pro Arg Gln Ser
                1          5          10

ccc cgc acc ctc cgt ccc gac ccg ggc ccc gcc atg tcc ttc ttc cgg      218
Pro Arg Thr Leu Arg Pro Asp Pro Gly Pro Ala Met Ser Phe Phe Arg
                15          20          25

cgg aaa gtg aaa ggc aaa gaa caa gag aag acc tca gat gtg aag tcc      266
Arg Lys Val Lys Gly Lys Glu Gln Glu Lys Thr Ser Asp Val Lys Ser
                30          35          40

att aaa gct tca ata tcc gta cat tcc cca caa aaa agc act aaa aat      314
Ile Lys Ala Ser Ile Ser Val His Ser Pro Gln Lys Ser Thr Lys Asn
                45          50          55

cat gcc ttg ctg gag gct gca gga cca agt cat gtt gca atc aat gcc      362
His Ala Leu Leu Glu Ala Ala Gly Pro Ser His Val Ala Ile Asn Ala
                60          65          70          75

att tct gcc aac atg gac tcc ttt tca agt agc agg aca gcc aca ctt      410
Ile Ser Ala Asn Met Asp Ser Phe Ser Ser Ser Arg Thr Ala Thr Leu
                80          85          90

aag aag cag cca agc cac atg gag gct gct cat ttt ggt gac ctg ggc      458
Lys Lys Gln Pro Ser His Met Glu Ala Ala His Phe Gly Asp Leu Gly
                95          100          105

aga tct tgt ctg gac tac cag act caa gag acc aaa tca agc ctt tct      506
Arg Ser Cys Leu Asp Tyr Gln Thr Gln Glu Thr Lys Ser Ser Leu Ser
                110          115          120

aag acc ctt gaa caa gtc ttg cac gac act att gtc ctc cct tac ttc      554
Lys Thr Leu Glu Gln Val Leu His Asp Thr Ile Val Leu Pro Tyr Phe
                125          130          135

att caa ttc atg gaa ctt cgg cga atg gag cat ttg gtg aaa ttt tgg      602
Ile Gln Phe Met Glu Leu Arg Arg Met Glu His Leu Val Lys Phe Trp
                140          145          150          155

tta gag gct gaa agt ttt cat tca aca act tgg tcg cga ata aga gca      650
Leu Glu Ala Glu Ser Phe His Ser Thr Thr Trp Ser Arg Ile Arg Ala
                160          165          170

cac agt cta aac aca atg aag cag agc tca ctg gct gag cct gtc tct      698
His Ser Leu Asn Thr Met Lys Gln Ser Ser Leu Ala Glu Pro Val Ser
                175          180          185

cca tct aaa aag cat gaa act aca gcg tct ttt tta act gat tct ctt      746
Pro Ser Lys Lys His Glu Thr Thr Ala Ser Phe Leu Thr Asp Ser Leu
                190          195          200

gat aag aga ttg gag gat tct ggc tca gca cag ttg ttt atg act cat      794

```

8/122

Asp	Lys	Arg	Leu	Glu	Asp	Ser	Gly	Ser	Ala	Gln	Leu	Phe	Met	Thr	His	
205						210					215					
tca	gaa	gga	att	gac	ctg	aat	aat	aga	act	aac	agc	act	cag	aat	cac	842
Ser	Glu	Gly	Ile	Asp	Leu	Asn	Asn	Arg	Thr	Asn	Ser	Thr	Gln	Asn	His	
220					225					230					235	
ttg	ctg	ctt	tcc	cag	gaa	tgt	gac	agt	gcc	cat	tct	ctc	cgt	ctt	gaa	890
Leu	Leu	Leu	Ser	Gln	Glu	Cys	Asp	Ser	Ala	His	Ser	Leu	Arg	Leu	Glu	
				240					245					250		
atg	gcc	aga	gca	gga	act	cac	caa	gtt	tcc	atg	gaa	acc	caa	gaa	tct	938
Met	Ala	Arg	Ala	Gly	Thr	His	Gln	Val	Ser	Met	Glu	Thr	Gln	Glu	Ser	
			255					260					265			
tcc	tct	aca	ctt	aca	gta	gcc	agt	aga	aat	agt	ccc	gct	tct	cca	cta	986
Ser	Ser	Thr	Leu	Thr	Val	Ala	Ser	Arg	Asn	Ser	Pro	Ala	Ser	Pro	Leu	
		270					275					280				
aaa	gaa	ttg	tca	gga	aaa	cta	atg	aaa	agt	ata	gaa	caa	gat	gca	gtg	1034
Lys	Glu	Leu	Ser	Gly	Lys	Leu	Met	Lys	Ser	Ile	Glu	Gln	Asp	Ala	Val	
	285					290					295					
aat	act	ttt	acc	aaa	tat	ata	tct	cca	gat	gct	gct	aaa	cca	ata	cca	1082
Asn	Thr	Phe	Thr	Lys	Tyr	Ile	Ser	Pro	Asp	Ala	Ala	Lys	Pro	Ile	Pro	
300					305					310					315	
att	aca	gaa	gca	atg	aga	aat	gac	atc	ata	gca	agg	att	tgt	gga	gaa	1130
Ile	Thr	Glu	Ala	Met	Arg	Asn	Asp	Ile	Ile	Ala	Arg	Ile	Cys	Gly	Glu	
				320					325					330		
gat	gga	cag	gtg	gat	ccc	aac	tgt	ttc	gtt	ttg	gca	cag	tcc	ata	gtc	1178
Asp	Gly	Gln	Val	Asp	Pro	Asn	Cys	Phe	Val	Leu	Ala	Gln	Ser	Ile	Val	
			335					340					345			
ttt	agt	gca	atg	gag	caa	gag	cac	ttt	agt	gag	ttt	ctg	cga	agt	cac	1226
Phe	Ser	Ala	Met	Glu	Gln	Glu	His	Phe	Ser	Glu	Phe	Leu	Arg	Ser	His	
		350					355					360				
cat	ttc	tgt	aaa	tac	cag	att	gaa	gtg	ctg	acc	agt	gga	act	gtt	tac	1274
His	Phe	Cys	Lys	Tyr	Gln	Ile	Glu	Val	Leu	Thr	Ser	Gly	Thr	Val	Tyr	
		365				370					375					
ctg	gct	gac	att	ctc	ttc	tgt	gag	tca	gcc	ctc	ttt	tat	ttc	tct	gag	1322
Leu	Ala	Asp	Ile	Leu	Phe	Cys	Glu	Ser	Ala	Leu	Phe	Tyr	Phe	Ser	Glu	
380					385					390					395	
tac	atg	gaa	aaa	gag	gat	gca	gtg	aat	atc	tta	caa	ttc	tgg	ttg	gca	1370
Tyr	Met	Glu	Lys	Glu	Asp	Ala	Val	Asn	Ile	Leu	Gln	Phe	Trp	Leu	Ala	
				400					405					410		
gca	gat	aac	ttc	cag	tct	cag	ctt	gct	gcc	aaa	aag	ggg	caa	tat	gat	1418
Ala	Asp	Asn	Phe	Gln	Ser	Gln	Leu	Ala	Ala	Lys	Lys	Gly	Gln	Tyr	Asp	
			415					420					425			
gga	cag	gag	gca	cag	aat	gat	gcc	atg	att	tta	tat	gac	aag	tac	ttc	1466
Gly	Gln	Glu	Ala	Gln	Asn	Asp	Ala	Met	Ile	Leu	Tyr	Asp	Lys	Tyr	Phe	
		430					435					440				
tcc	ctc	caa	gcc	aca	cat	cct	ctt	gga	ttt	gat	gat	gtt	gta	cga	tta	1514

9/122

Ser 445	Leu	Gln	Ala	Thr	His	Pro 450	Leu	Gly	Phe	Asp 455	Asp	Val	Val	Arg	Leu		
gaa	att	gaa	tcc	aat	atc	tgc	agg	gaa	ggg	ggg	cca	ctc	ccc	aac	tgt	1562	
Glu	Ile	Glu	Ser	Asn	Ile	Cys	Arg	Glu	Gly	Gly	Pro	Leu	Pro	Asn	Cys		
460					465					470					475		
ttc	aca	act	cca	tta	cgt	cag	gcc	tgg	aca	acc	atg	gag	aag	gtc	ttt	1610	
Phe	Thr	Thr	Pro	Leu	Arg	Gln	Ala	Trp	Thr	Thr	Met	Glu	Lys	Val	Phe		
				480					485					490			
ttg	cct	ggc	ttt	ctg	tcc	agc	aat	ctt	tat	tat	aaa	tat	ttg	aat	gat	1658	
Leu	Pro	Gly	Phe	Leu	Ser	Ser	Asn	Leu	Tyr	Tyr	Lys	Tyr	Leu	Asn	Asp		
			495					500					505				
ctc	atc	cat	tgc	gtt	cga	gga	gat	gaa	ttt	ctg	ggc	ggg	aac	gtg	tgc	1706	
Leu	Ile	His	Ser	Val	Arg	Gly	Asp	Glu	Phe	Leu	Gly	Gly	Asn	Val	Ser		
		510					515					520					
ccg	act	gct	cct	ggc	tct	gtt	ggc	cct	cct	gat	gag	tct	cac	cca	ggg	1754	
Pro	Thr	Ala	Pro	Gly	Ser	Val	Gly	Pro	Pro	Asp	Glu	Ser	His	Pro	Gly		
	525					530					535						
agt	tct	gac	agc	tct	gcg	tct	cag	tcc	agt	gtg	aaa	aaa	gcc	agt	att	1802	
Ser	Ser	Asp	Ser	Ser	Ala	Ser	Gln	Ser	Ser	Val	Lys	Lys	Ala	Ser	Ile		
540					545					550					555		
aaa	ata	ctg	aaa	aat	ttt	gat	gaa	gcg	ata	att	gtg	gat	gcg	gca	agt	1850	
Lys	Ile	Leu	Lys	Asn	Phe	Asp	Glu	Ala	Ile	Ile	Val	Asp	Ala	Ala	Ser		
				560					565					570			
ctg	gat	cca	gaa	tct	tta	tat	caa	cgg	aca	tat	gcc	ggg	aag	atg	aca	1898	
Leu	Asp	Pro	Glu	Ser	Leu	Tyr	Gln	Arg	Thr	Tyr	Ala	Gly	Lys	Met	Thr		
			575					580					585				
ttt	gga	aga	gtg	agt	gac	ttg	ggg	caa	ttc	atc	cgg	gaa	tct	gag	cct	1946	
Phe	Gly	Arg	Val	Ser	Asp	Leu	Gly	Gln	Phe	Ile	Arg	Glu	Ser	Glu	Pro		
	590						595					600					
gaa	cct	gat	gta	agg	aaa	tca	aaa	gga	tcc	atg	ttc	tca	caa	gct	atg	1994	
Glu	Pro	Asp	Val	Arg	Lys	Ser	Lys	Gly	Ser	Met	Phe	Ser	Gln	Ala	Met		
	605					610					615						
aag	aaa	tgg	gtg	caa	gga	aat	act	gat	gag	gcc	cag	gaa	gag	cta	gct	2042	
Lys	Lys	Trp	Val	Gln	Gly	Asn	Thr	Asp	Glu	Ala	Gln	Glu	Glu	Leu	Ala		
620					625					630					635		
tgg	aag	att	gct	aaa	atg	ata	gtc	agt	gac	att	atg	cag	cag	gct	cag	2090	
Trp	Lys	Ile	Ala	Lys	Met	Ile	Val	Ser	Asp	Ile	Met	Gln	Gln	Ala	Gln		
				640					645					650			
tat	gat	caa	ccg	tta	gag	aaa	tct	aca	aag	tta	tga	ctcaaaaactt				2136	
Tyr	Asp	Gln	Pro	Leu	Glu	Lys	Ser	Thr	Lys	Leu	*						
				655				660									
gagataaagg	aaatctgctt	gtgaaaaata	agagaacttt	tttcccttgg	ttggattctt											2196	
caacacagcc	aatgaaaaca	gcactatatt	tctgatctgt	cactgttgtt	tccagggaga											2256	
gaatggggag	acaatcctag	gacttccacc	ctaatagcagt	tacctgtagg	gcataattgg											2316	
atggcacatg	atgtttcaca	cagtgaggag	tctttaaagg	ttaccaa												2363	

10/122

<210> 32  
 <211> 662  
 <212> PRT  
 <213> Homo Sapien

<400> 32  
 Met Arg Gly Ala Gly Pro Ser Pro Arg Gln Ser Pro Arg Thr Leu Arg  
 1 5 10 15  
 Pro Asp Pro Gly Pro Ala Met Ser Phe Arg Arg Lys Val Lys Gly  
 20 25 30  
 Lys Glu Gln Glu Lys Thr Ser Asp Val Lys Ser Ile Lys Ala Ser Ile  
 35 40 45  
 Ser Val His Ser Pro Gln Lys Ser Thr Lys Asn His Ala Leu Leu Glu  
 50 55 60  
 Ala Ala Gly Pro Ser His Val Ala Ile Asn Ala Ile Ser Ala Asn Met  
 65 70 75 80  
 Asp Ser Phe Ser Ser Ser Arg Thr Ala Thr Leu Lys Lys Gln Pro Ser  
 85 90 95  
 His Met Glu Ala Ala His Phe Gly Asp Leu Gly Arg Ser Cys Leu Asp  
 100 105 110  
 Tyr Gln Thr Gln Glu Thr Lys Ser Ser Leu Ser Lys Thr Leu Glu Gln  
 115 120 125  
 Val Leu His Asp Thr Ile Val Leu Pro Tyr Phe Ile Gln Phe Met Glu  
 130 135 140  
 Leu Arg Arg Met Glu His Leu Val Lys Phe Trp Leu Glu Ala Glu Ser  
 145 150 155 160  
 Phe His Ser Thr Thr Trp Ser Arg Ile Arg Ala His Ser Leu Asn Thr  
 165 170 175  
 Met Lys Gln Ser Ser Leu Ala Glu Pro Val Ser Pro Ser Lys Lys His  
 180 185 190  
 Glu Thr Thr Ala Ser Phe Leu Thr Asp Ser Leu Asp Lys Arg Leu Glu  
 195 200 205  
 Asp Ser Gly Ser Ala Gln Leu Phe Met Thr His Ser Glu Gly Ile Asp  
 210 215 220  
 Leu Asn Asn Arg Thr Asn Ser Thr Gln Asn His Leu Leu Leu Ser Gln  
 225 230 235 240  
 Glu Cys Asp Ser Ala His Ser Leu Arg Leu Glu Met Ala Arg Ala Gly  
 245 250 255  
 Thr His Gln Val Ser Met Glu Thr Gln Glu Ser Ser Ser Thr Leu Thr  
 260 265 270  
 Val Ala Ser Arg Asn Ser Pro Ala Ser Pro Leu Lys Glu Leu Ser Gly  
 275 280 285  
 Lys Leu Met Lys Ser Ile Glu Gln Asp Ala Val Asn Thr Phe Thr Lys  
 290 295 300  
 Tyr Ile Ser Pro Asp Ala Ala Lys Pro Ile Pro Ile Thr Glu Ala Met  
 305 310 315 320  
 Arg Asn Asp Ile Ile Ala Arg Ile Cys Gly Glu Asp Gly Gln Val Asp  
 325 330 335  
 Pro Asn Cys Phe Val Leu Ala Gln Ser Ile Val Phe Ser Ala Met Glu  
 340 345 350  
 Gln Glu His Phe Ser Glu Phe Leu Arg Ser His His Phe Cys Lys Tyr  
 355 360 365  
 Gln Ile Glu Val Leu Thr Ser Gly Thr Val Tyr Leu Ala Asp Ile Leu  
 370 375 380  
 Phe Cys Glu Ser Ala Leu Phe Tyr Phe Ser Glu Tyr Met Glu Lys Glu  
 385 390 395 400  
 Asp Ala Val Asn Ile Leu Gln Phe Trp Leu Ala Ala Asp Asn Phe Gln  
 405 410 415  
 Ser Gln Leu Ala Ala Lys Lys Gly Gln Tyr Asp Gly Gln Glu Ala Gln  
 420 425 430

11/122

Asn Asp Ala Met Ile Leu Tyr Asp Lys Tyr Phe Ser Leu Gln Ala Thr  
           435                                  440                  445  
 His Pro Leu Gly Phe Asp Asp Val Val Arg Leu Glu Ile Glu Ser Asn  
           450                                  455                  460  
 Ile Cys Arg Glu Gly Gly Pro Leu Pro Asn Cys Phe Thr Thr Pro Leu  
           465                                  470                  475                  480  
 Arg Gln Ala Trp Thr Thr Met Glu Lys Val Phe Leu Pro Gly Phe Leu  
                                   485                                  490                  495  
 Ser Ser Asn Leu Tyr Tyr Lys Tyr Leu Asn Asp Leu Ile His Ser Val  
                                   500                                  505                  510  
 Arg Gly Asp Glu Phe Leu Gly Gly Asn Val Ser Pro Thr Ala Pro Gly  
                                   515                                  520                  525  
 Ser Val Gly Pro Pro Asp Glu Ser His Pro Gly Ser Ser Asp Ser Ser  
                                   530                                  535                  540  
 Ala Ser Gln Ser Ser Val Lys Lys Ala Ser Ile Lys Ile Leu Lys Asn  
                                   545                                  550                  555                  560  
 Phe Asp Glu Ala Ile Ile Val Asp Ala Ala Ser Leu Asp Pro Glu Ser  
                                   565                                  570                  575  
 Leu Tyr Gln Arg Thr Tyr Ala Gly Lys Met Thr Phe Gly Arg Val Ser  
                                   580                                  585                  590  
 Asp Leu Gly Gln Phe Ile Arg Glu Ser Glu Pro Glu Pro Asp Val Arg  
                                   595                                  600                  605  
 Lys Ser Lys Gly Ser Met Phe Ser Gln Ala Met Lys Lys Trp Val Gln  
                                   610                                  615                  620  
 Gly Asn Thr Asp Glu Ala Gln Glu Glu Leu Ala Trp Lys Ile Ala Lys  
                                   625                                  630                  635                  640  
 Met Ile Val Ser Asp Ile Met Gln Gln Ala Gln Tyr Asp Gln Pro Leu  
                                   645                                  650                  655  
 Glu Lys Ser Thr Lys Leu  
                                   660

&lt;210&gt; 33

&lt;211&gt; 2363

&lt;212&gt; DNA

&lt;213&gt; Homo Sapien

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (138)...(2126)

&lt;223&gt; AKAP-10-5

&lt;221&gt; allele

&lt;222&gt; 2073

&lt;223&gt; Single Nucleotide Polymorphism: A to G

&lt;400&gt; 33

gcggcttggt gataatatgg cggctggagc tgccctgggca tcccaggagg gcggtggggc 60  
 ccactcccgg aagaagggtc cttttcgcg ctagtgcagc ggcccctctg gacccggaag 120  
 tccggggccgg ttgctga atg agg gga gcc ggg ccc tcc ccg cgc cag tcc 170  
                   Met Arg Gly Ala Gly Pro Ser Pro Arg Gln Ser  
                                   1                                  5                                  10  
 ccc cgc acc ctc cgt ccc gac ccg ggc ccc gcc atg tcc ttc ttc cgg 218  
 Pro Arg Thr Leu Arg Pro Asp Pro Gly Pro Ala Met Ser Phe Phe Arg  
                                   15                                  20                                  25  
 cgg aaa gtg aaa ggc aaa gaa caa gag aag acc tca gat gtg aag tcc 266  
 Arg Lys Val Lys Gly Lys Glu Gln Glu Lys Thr Ser Asp Val Lys Ser  
                                   30                                  35                                  40

12/122

att aaa gct tca ata tcc gta cat tcc cca caa aaa agc act aaa aat	314
Ile Lys Ala Ser Ile Ser Val His Ser Pro Gln Lys Ser Thr Lys Asn	
45 50 55	
cat gcc ttg ctg gag gct gca gga cca agt cat gtt gca atc aat gcc	362
His Ala Leu Leu Glu Ala Ala Gly Pro Ser His Val Ala Ile Asn Ala	
60 65 70 75	
att tct gcc aac atg gac tcc ttt tca agt agc agg aca gcc aca ctt	410
Ile Ser Ala Asn Met Asp Ser Phe Ser Ser Ser Arg Thr Ala Thr Leu	
80 85 90	
aag aag cag cca agc cac atg gag gct gct cat ttt ggt gac ctg ggc	458
Lys Lys Gln Pro Ser His Met Glu Ala Ala His Phe Gly Asp Leu Gly	
95 100 105	
aga tct tgt ctg gac tac cag act caa gag acc aaa tca agc ctt tct	506
Arg Ser Cys Leu Asp Tyr Gln Thr Gln Glu Thr Lys Ser Ser Leu Ser	
110 115 120	
aag aac ctt gaa caa gtc ttg cac gac act att gtc ctc cct tac ttc	554
Lys Thr Leu Glu Gln Val Leu His Asp Thr Ile Val Leu Pro Tyr Phe	
125 130 135	
att caa ttc atg gaa ctt cgg cga atg gag cat ttg gtg aaa ttt tgg	602
Ile Gln Phe Met Glu Leu Arg Arg Met Glu His Leu Val Lys Phe Trp	
140 145 150 155	
tta gag gct gaa agt ttt cat tca aca act tgg tcg cga ata aga gca	650
Leu Glu Ala Glu Ser Phe His Ser Thr Thr Trp Ser Arg Ile Arg Ala	
160 165 170	
cac agt cta aac aca atg aag cag agc tca ctg gct gag cct gtc tct	698
His Ser Leu Asn Thr Met Lys Gln Ser Ser Leu Ala Glu Pro Val Ser	
175 180 185	
cca tct aaa aag cat gaa act aca gcg tct ttt tta act gat tct ctt	746
Pro Ser Lys Lys His Glu Thr Thr Ala Ser Phe Leu Thr Asp Ser Leu	
190 195 200	
gat aag aga ttg gag gat tct ggc tca gca cag ttg ttt atg act cat	794
Asp Lys Arg Leu Glu Asp Ser Gly Ser Ala Gln Leu Phe Met Thr His	
205 210 215	
tca gaa gga att gac ctg aat aat aga act aac agc act cag aat cac	842
Ser Glu Gly Ile Asp Leu Asn Asn Arg Thr Asn Ser Thr Gln Asn His	
220 225 230 235	
ttg ctg ctt tcc cag gaa tgt gac agt gcc cat tct ctc cgt ctt gaa	890
Leu Leu Leu Ser Gln Glu Cys Asp Ser Ala His Ser Leu Arg Leu Glu	
240 245 250	
atg gcc aga gca gga act cac caa gtt tcc atg gaa acc caa gaa tct	938
Met Ala Arg Ala Gly Thr His Gln Val Ser Met Glu Thr Gln Glu Ser	
255 260 265	
tcc tct aca ctt aca gta gcc agt aga aat agt ccc gct tct cca cta	986
Ser Ser Thr Leu Thr Val Ala Ser Arg Asn Ser Pro Ala Ser Pro Leu	
270 275 280	

13/122

aaa gaa ttg tca gga aaa cta atg aaa agt ata gaa caa gat gca gtg Lys Glu Leu Ser Gly Lys Leu Met Lys Ser Ile Glu Gln Asp Ala Val 285 290 295	1034
aat act ttt acc aaa tat ata tct cca gat gct gct aaa cca ata cca Asn Thr Phe Thr Lys Tyr Ile Ser Pro Asp Ala Lys Pro Ile Pro 300 305 310 315	1082
att aca gaa gca atg aga aat gac atc ata gca agg att tgt gga gaa Ile Thr Glu Ala Met Arg Asn Asp Ile Ile Ala Arg Ile Cys Gly Glu 320 325 330	1130
gat gga cag gtg gat ccc aac tgt ttc gtt ttg gca cag tcc ata gtc Asp Gly Gln Val Asp Pro Asn Cys Phe Val Leu Ala Gln Ser Ile Val 335 340 345	1178
ttt agt gca atg gag caa gag cac ttt agt gag ttt ctg cga agt cac Phe Ser Ala Met Glu Gln Glu His Phe Ser Glu Phe Leu Arg Ser His 350 355 360	1226
cat ttc tgt aaa tac cag att gaa gtg ctg acc agt gga act gtt tac His Phe Cys Lys Tyr Gln Ile Glu Val Leu Thr Ser Gly Thr Val Tyr 365 370 375	1274
ctg gct gac att ctc ttc tgt gag tca gcc ctc ttt tat ttc tct gag Leu Ala Asp Ile Leu Phe Cys Glu Ser Ala Leu Phe Tyr Phe Ser Glu 380 385 390 395	1322
tac atg gaa aaa gag gat gca gtg aat atc tta caa ttc tgg ttg gca Tyr Met Glu Lys Glu Asp Ala Val Asn Ile Leu Gln Phe Trp Leu Ala 400 405 410	1370
gca gat aac ttc cag tct cag ctt gct gcc aaa aag ggg caa tat gat Ala Asp Asn Phe Gln Ser Gln Leu Ala Ala Lys Lys Gly Gln Tyr Asp 415 420 425	1418
gga cag gag gca cag aat gat gcc atg att tta tat gac aag tac ttc Gly Gln Glu Ala Gln Asn Asp Ala Met Ile Leu Tyr Asp Lys Tyr Phe 430 435 440	1466
tcc ctc caa gcc aca cat cct ctt gga ttt gat gat gtt gta cga tta Ser Leu Gln Ala Thr His Pro Leu Gly Phe Asp Asp Val Val Arg Leu 445 450 455	1514
gaa att gaa tcc aat atc tgc agg gaa ggt ggg cca ctc ccc aac tgt Glu Ile Glu Ser Asn Ile Cys Arg Glu Gly Pro Leu Pro Asn Cys 460 465 470 475	1562
ttc aca act cca tta cgt cag gcc tgg aca acc atg gag aag gtc ttt Phe Thr Thr Pro Leu Arg Gln Ala Trp Thr Thr Met Glu Lys Val Phe 480 485 490	1610
ttg cct ggc ttt ctg tcc agc aat ctt tat tat aaa tat ttg aat gat Leu Pro Gly Phe Leu Ser Ser Asn Leu Tyr Tyr Lys Tyr Leu Asn Asp 495 500 505	1658
ctc atc cat tcg gtt cga gga gat gaa ttt ctg ggc ggg aac gtg tcg Leu Ile His Ser Val Arg Gly Asp Glu Phe Leu Gly Gly Asn Val Ser 510 515 520	1706

14/122

ccg act gct cct ggc tct gtt ggc cct cct gat gag tct cac cca ggg	1754
Pro Thr Ala Pro Gly Ser Val Gly Pro Pro Asp Glu Ser His Pro Gly	
525 530 535	
agt tct gac agc tct gcg tct cag tcc agt gtg aaa aaa gcc agt att	1802
Ser Ser Asp Ser Ser Ala Ser Gln Ser Ser Val Lys Lys Ala Ser Ile	
540 545 550 555	
aaa ata ctg aaa aat ttt gat gaa gcg ata att gtg gat gcg gca agt	1850
Lys Ile Leu Lys Asn Phe Asp Glu Ala Ile Val Asp Ala Ala Ser	
560 565 570	
ctg gat cca gaa tct tta tat caa cgg aca tat gcc ggg aag atg aca	1898
Leu Asp Pro Glu Ser Leu Tyr Gln Arg Thr Tyr Ala Gly Lys Met Thr	
575 580 585	
ttt gga aga gtg agt gac ttg ggg caa ttc atc cgg gaa tct gag cct	1946
Phe Gly Arg Val Ser Asp Leu Gly Gln Phe Ile Arg Glu Ser Glu Pro	
590 595 600	
gaa cct gat gta agg aaa tca aaa gga tcc atg ttc tca caa gct atg	1994
Glu Pro Asp Val Arg Lys Ser Lys Gly Ser Met Phe Ser Gln Ala Met	
605 610 615	
aag aaa tgg gtg caa gga aat act gat gag gcc cag gaa gag cta gct	2042
Lys Lys Trp Val Gln Gly Asn Thr Asp Glu Ala Gln Glu Glu Leu Ala	
620 625 630 635	
tgg aag att gct aaa atg ata gtc agt gac gtt atg cag cag gct cag	2090
Trp Lys Ile Ala Lys Met Ile Val Ser Asp Val Met Gln Gln Ala Gln	
640 645 650	
tat gat caa ccg tta gag aaa tct aca aag tta tga ctcaaaactt	2136
Tyr Asp Gln Pro Leu Glu Lys Ser Thr Lys Leu *	
655 660	
gagataaagg aaatctgctt gtgaaaaata agagaacttt tttcccttggt ttggattctt	2196
caacacagcc aatgaaaaca gcactatatt tctgatctgt cactggtggt tccagggaga	2256
gaatggggag acaatcctag gacttcacc ctaatgcagt tacctgtagg gcataattgg	2316
atggcacatg atgtttcaca cagtgaggag tctttaaagg ttaccaa	2363

&lt;210&gt; 34

&lt;211&gt; 662

&lt;212&gt; PRT

&lt;213&gt; Homo Sapien

&lt;400&gt; 34

Met Arg Gly Ala Gly Pro Ser Pro Arg Gln Ser Pro Arg Thr Leu Arg	
1 5 10 15	
Pro Asp Pro Gly Pro Ala Met Ser Phe Arg Arg Lys Val Lys Gly	
20 25 30	
Lys Glu Gln Glu Lys Thr Ser Asp Val Lys Ser Ile Lys Ala Ser Ile	
35 40 45	
Ser Val His Ser Pro Gln Lys Ser Thr Lys Asn His Ala Leu Leu Glu	
50 55 60	
Ala Ala Gly Pro Ser His Val Ala Ile Asn Ala Ile Ser Ala Asn Met	
65 70 75 80	
Asp Ser Phe Ser Ser Ser Arg Thr Ala Thr Leu Lys Lys Gln Pro Ser	
85 90 95	
His Met Glu Ala Ala His Phe Gly Asp Leu Gly Arg Ser Cys Leu Asp	

				100						105					110				
Tyr	Gln	Thr	Gln	Glu	Thr	Lys	Ser	Ser	Leu	Ser	Lys	Thr	Leu	Glu	Gln				
		115					120					125							
Val	Leu	His	Asp	Thr	Ile	Val	Leu	Pro	Tyr	Phe	Ile	Gln	Phe	Met	Glu				
	130					135					140								
Leu	Arg	Arg	Met	Glu	His	Leu	Val	Lys	Phe	Trp	Leu	Glu	Ala	Glu	Ser				
145					150					155					160				
Phe	His	Ser	Thr	Thr	Trp	Ser	Arg	Ile	Arg	Ala	His	Ser	Leu	Asn	Thr				
				165					170					175					
Met	Lys	Gln	Ser	Ser	Leu	Ala	Glu	Pro	Val	Ser	Pro	Ser	Lys	Lys	His				
			180					185					190						
Glu	Thr	Thr	Ala	Ser	Phe	Leu	Thr	Asp	Ser	Leu	Asp	Lys	Arg	Leu	Glu				
		195					200					205							
Asp	Ser	Gly	Ser	Ala	Gln	Leu	Phe	Met	Thr	His	Ser	Glu	Gly	Ile	Asp				
	210					215						220							
Leu	Asn	Asn	Arg	Thr	Asn	Ser	Thr	Gln	Asn	His	Leu	Leu	Leu	Ser	Gln				
225					230					235					240				
Glu	Cys	Asp	Ser	Ala	His	Ser	Leu	Arg	Leu	Glu	Met	Ala	Arg	Ala	Gly				
				245					250					255					
Thr	His	Gln	Val	Ser	Met	Glu	Thr	Gln	Glu	Ser	Ser	Ser	Thr	Leu	Thr				
			260					265						270					
Val	Ala	Ser	Arg	Asn	Ser	Pro	Ala	Ser	Pro	Leu	Lys	Glu	Leu	Ser	Gly				
		275				280						285							
Lys	Leu	Met	Lys	Ser	Ile	Glu	Gln	Asp	Ala	Val	Asn	Thr	Phe	Thr	Lys				
	290					295					300								
Tyr	Ile	Ser	Pro	Asp	Ala	Ala	Lys	Pro	Ile	Pro	Ile	Thr	Glu	Ala	Met				
305					310					315					320				
Arg	Asn	Asp	Ile	Ile	Ala	Arg	Ile	Cys	Gly	Glu	Asp	Gly	Gln	Val	Asp				
				325					330					335					
Pro	Asn	Cys	Phe	Val	Leu	Ala	Gln	Ser	Ile	Val	Phe	Ser	Ala	Met	Glu				
			340					345					350						
Gln	Glu	His	Phe	Ser	Glu	Phe	Leu	Arg	Ser	His	His	Phe	Cys	Lys	Tyr				
		355					360					365							
Gln	Ile	Glu	Val	Leu	Thr	Ser	Gly	Thr	Val	Tyr	Leu	Ala	Asp	Ile	Leu				
	370					375					380								
Phe	Cys	Glu	Ser	Ala	Leu	Phe	Tyr	Phe	Ser	Glu	Tyr	Met	Glu	Lys	Glu				
385					390														

16/122

580	585	590
Asp Leu Gly Gln Phe Ile Arg Glu	Ser Glu Pro Glu Pro Asp Val Arg	
595	600	605
Lys Ser Lys Gly Ser Met Phe Ser	Gln Ala Met Lys Lys Trp Val Gln	
610	615	620
Gly Asn Thr Asp Glu Ala Gln Glu	Glu Leu Ala Trp Lys Ile Ala Lys	
625	630	635
Met Ile Val Ser Asp Val Met Gln	Gln Ala Gln Tyr Asp Gln Pro Leu	
645	650	655
Glu Lys Ser Thr Lys Leu		
660		

<210> 35  
 <211> 162025  
 <212> DNA  
 <213> Homo Sapien

. 300>  
 . 308> GenBank AC005730  
 . 309> 1998-10-22

<400> 35

gaattcctat	ttcaaaagaa	acaaatgggc	caagtatggt	ggctcatacc	tgtaatccca	60
gcactttggg	aggccgaggt	gagtggttca	cttgaggtca	ggagttccag	gccagtctgg	120
ccaacatggt	gaaacactgt	ctctactaaa	aatacaaaaa	ttagccgggc	gtgggtggcgg	180
gcacctgtaa	tcccagctac	tcaggaggct	gaggcaggag	aattgcttga	acctgggaga	240
tggagggtgc	agtgagccga	gatcgcgcca	ctgctctcca	gcctgggtgg	cagagtggaga	300
ctctgtctca	aaaagaaaaca	aagaaataaa	tgaaacaatt	ttgttcacat	atatttcaca	360
aatttgaaat	gttaaaggta	ttatgggtcac	tgatatcctg	tttcattctt	tatataatca	420
ttaagtttga	aatgtatact	tgcactacta	acacagtagt	taatcttagt	cctacaagtt	480
actgctttta	cacaatatat	tttcgtaata	tgtatgcact	gggtgtttatg	tacgtgttta	540
tgtttatatc	tgttaaaatt	agcagtttcc	atctttttct	atattgtacc	atcacatcag	600
ttcagaagga	ttgacagagc	aaaatgattt	gatgaagtat	aaaagtcaca	tggtgagtgg	660
cataaataca	actctgaaca	attaggaggc	tcactattga	ctggaactaa	actgcaagcc	720
agaaagacac	atatcctata	tgtcaagaga	tgtaccaccc	aggcagttaa	agaagggaag	780
tacacataga	aagcacaatg	gtgaataatt	aaaaaattgg	aatttatcag	acactggatt	840
catttgctcc	taaagtccaga	gtcctctatt	gtttttttgt	ttttgtgggt	ttctttttaa	900
atttttttat	tttttgtaga	gtcggagtct	cactgtgtta	ccccgggtgg	tctagaactc	960
ctggcctcaa	acaaacctcc	tgcctcagct	tcccaaagca	ttgggattac	agacatgagc	1020
cactgagccc	agcccagacg	cttttagcatt	tatgaagctt	ctgaaatagt	tgtagaaacc	1080
gcataagctt	tccatgtcac	tttcaaaagt	tgatggcttc	tttagtaaac	caaccaagtt	1140
attcctcaag	ggcaaaataa	cattttctcag	tgcaaaactg	atgcacttca	ttaccaaaaag	1200
gaaaagacca	caactataga	ggcgtcattg	aaagctgcac	tcttcagagg	ccaaaaaaaa	1260
aggtacaaac	acatactaata	ggaacattct	ttagaagagc	cccaaagtta	atgataaaca	1320
ttttcatcaa	agagaaaaga	gaacaagggtg	ttagcaaatt	cctctatcaa	ataacactaa	1380
acatcaagga	acatcaatgg	catgccatgt	ggaagaggaa	gtgctagctc	atgtacaaac	1440
cagtagataa	tttcaacttg	ctgccgaatg	aaacctcttt	gcaagggtatg	aatcagcact	1500
tctcatgttt	gttttgcttt	gttttgtttt	gttttttagag	acaggccctt	gctctgtcac	1560
acaggcttga	gtgcagtggc	acgatcagag	ctcactgcaa	cctgaaactc	ctggggtcaa	1620
gggatccctc	tgccttagcc	tccaagtag	ctgggactac	aggcccacca	tgcccagcta	1680
atttttttaa	ttttctatag	agatgggatc	tcactagcac	ctttcatgtt	tgatgttcat	1740
atacaacgac	caaggtagaa	tgtggaaaag	gggtctcagg	atctaaagtg	aaggaggacc	1800
agaaagaaaa	gggggttgcta	catagagtag	aagaagttgc	acttcatgcc	agtctacaac	1860
actgctgttt	tcttcagagc	agagttgatg	atctaaatca	gggggtcccca	acccccagtt	1920
catagcctgt	taggaaccgg	gccacacagc	aggaggtgag	caataggcaa	gcgagcatta	1980
ccacctgggc	ttcacctccc	gtcagatcag	tgatgtcatt	agattctcat	aggaccatga	2040
accctattgt	gaactgagca	tgcaagggat	gtaggttttc	cgctctttat	gagactctaa	2100
tgccggaaga	tctgtcactg	tcttccatca	ccctgagatg	ggaacatcta	gttgaggagg	2160
aacaacctca	ggggtcccatt	tgattctata	ttacagttag	ttgtatcatt	atttcattct	2220
atattacaat	gtaataataa	tagaaataaa	ggcacaatag	gccaggcgtg	gtgggtcaca	2280

17/122

cctgtaatcc	cagcacttcg	ggaggccaag	gcaggcggat	cacgaggtca	ggagatcgag	2340
accatcctgg	ctaaaacggt	gaaacccccg	ctactaaaaa	ttcaaaaaaa	aattagccgg	2400
gtgtggtgg	gggcacctgt	agtcccagct	actcgagagg	ctgaggcagg	agaatggtgt	2460
gaacctggga	ggcagagctt	gaggttaagcc	gagatcacgc	cactgcactc	cagcctgggc	2520
gacagagcga	tactctgtct	caaaaaaaaaa	aaaaaaaaaa	aaagaaataa	agtgaacaat	2580
aaatgtaatg	tggctgaatc	attccaaaaa	aatcccccca	ccccagttca	cggaaaaaatt	2640
ctcccacaaa	accagtcctt	ggtgccaaaa	agggtgggga	ccgctaattc	aaataatcta	2700
atcttcattc	aatgctaaaa	aatgaataaaa	ctttttttta	aatacacggt	ctcactttgt	2760
tgcccagggt	ggagtacggt	ggcatgatca	cagctcactg	tagcctcaat	cacccaggcc	2820
ccagcgatcc	tcccacctaa	acttcctgag	tagctgggac	tacaggcacg	caccaccatg	2880
cccagctaata	ttttaaattt	tttatagaga	tgggggtctc	accatgttgc	ccagactggg	2940
ctcaaaccct	gggctcaagt	gatcctccct	caaactcctg	gactcaagtg	atcctccttc	3000
cttggcctcc	caaagtgtct	ggattacaag	catgagccac	tgtacccagc	tggataaaca	3060
ttttaagtct	cactacagtc	atggacaatc	aggcttttca	acatgcagta	tggacagtga	3120
gtcccagggt	ctgcttttcc	atactgaaat	acatgtgata	ctaaggagaa	agggtgctcg	3180
aaggatattt	aaaatgaaga	atatttaaaa	tgaggaaaaa	actgtttctt	catgactttg	3240
ataaggctga	taaagacccat	ttctgtgatc	tcaggtgatt	cactcaagta	gtatatttca	3300
gtaatcatta	tctggaacag	cctgaatctt	aaccaaataa	ccatgatatt	ttaatgtctg	3360
tatgatacct	tgatgatatg	accaaactgc	aatgtaggca	gctaaatctc	cacgagtttg	3420
acttccccga	gagttgacag	ttttcttcac	aaattaaaga	aatatatttt	ttgatacatg	3480
attggcatat	ttaaaaacta	cactgaaatg	ctgcaaaatg	atataaagaa	acattttcca	3540
gaatcaaatg	caatcaaaga	gtggattagg	aatctactca	ccattatcaa	ctaaatagaa	3600
acacttggac	tgggtgtggt	ggctcacatc	tgtaatctca	gcactttggg	aggccaaggc	3660
agggtggattg	cttgaggcca	ggagctcaag	accagcctga	gcaacatagc	aaaactctgt	3720
ctctacaaaa	aaaaaaaaaa	attaaccagg	catggtggca	gatgcttgta	atcccagcta	3780
ctctggaagc	tgaagtagga	ggactgcttg	agcccaggag	atcaagactg	cagtgaagccg	3840
tggctcatgct	gcgccacagc	ctgagtgaac	gagagagacc	ctgtctcaaa	aacaaaaaca	3900
aacaaaaaac	acttaacctt	cctgtttttt	gctgttgttg	ttgttgtttg	tttgttttga	3960
gatggagtct	cactctgttg	cccaggctgg	agtgcagtgg	cgtgatcttg	gctcactgca	4020
agctctgcct	cccgggttca	cgccattctc	ctgcctcagc	ctcccagagta	gctgggacta	4080
taggcgcccc	ccaccacgcc	cggctacttt	tttgcatttt	tagtagagat	gggggtttcac	4140
cgtgttagcc	aggatgggtc	tgatctcctg	accctgtgat	ccacctgcct	cggcctccca	4200
aagtgtctgg	attacaggca	tgagccaccg	caccggcca	acctttctgt	tttttagttt	4260
gatatgcttg	ttactcagc	agctgaaaga	atgctgaaag	tggccttcag	taaaaaaatt	4320
tcaatagaat	ctctacatcc	atattttaatc	tgaatgcata	tccagattga	tcagttagag	4380
caaaaacact	catcatcatt	cctgatgacc	tctaattctg	gtttcggctt	tctatttcaa	4440
tggaaacaga	ataagggaag	aaatgggaag	gctctggaaa	tttgtcctgg	gctatagata	4500
ctatcaaaga	tcaccaacaa	taagatctct	cctataaata	taaaaacaagt	ataattaatt	4560
ttttaattat	ttttttctct	tcagaggatt	ttatttcaag	ataaaacata	acttctaccc	4620
atactattga	ttccaaaggt	tagaaaaagt	gtttttcctc	atcttatcct	tcaaagagggt	4680
cacagcaatg	caaacatcta	taaaatgcct	ctgcataatt	gtcagaagct	atagtccaga	4740
aatcattgaa	aatgcttttc	catttttaagc	ttaggtgagg	tgtcttagga	aacctctatg	4800
acaacttact	ctattttattg	ggaggtaaac	tcccagactc	tcccagggtc	tctgtatttg	4860
atctcatttt	ttaggcttcc	taatcccttg	aagcacatc	gaaaaagccc	tggatctctt	4920
ttctgcacat	atcatcgcg	aattcattcg	gcttccagca	agctgacact	ccatgataca	4980
agcggcctcg	cccttctccg	gacgccagtc	cttgctggcg	ttagctagga	tgagggggtt	5040
gctgggcttc	agtgcaggct	tctgcggtt	cccaagccgc	accagggtgg	ctcacaggct	5100
ggatgtcacc	attgcacact	gagctcctgg	caggctgtac	caatttttta	attattttaat	5160
atctattttt	aaaattatgg	tgaatatttt	ggtattctgc	tctaaaatag	gcccataaat	5220
gcacagcaga	tatctcttgg	aacccacagc	tttccactgg	aagaactaag	tatttttctt	5280
ttaaagatgc	tactaagtct	ctgaaaagtc	cagatcctct	acctctttcc	atcccaaact	5340
aagacttgga	atztatgaga	gatctagcta	acagaaatcc	cagacacatc	attggttctt	5400
cccagagtgc	agtcctccta	aagagggtca	gccctaagca	ggcccctgca	ccaggagggt	5460
gggtctgaga	cccacatagc	acttcccaac	gtgcatgctc	cagagaggca	ctgaaacagc	5520
tgagcacaag	cctgcaagcc	tggagaactc	tcacagtcag	aacggagggg	gcccagtggt	5580
actaacataa	agagaaaagg	gaacacagag	aaatggatgg	caccaacaac	cagcaaagcc	5640
ttcatggcca	atgaaagcat	cagtgcgggg	gccagaaccc	tcatcccaa	agactcttca	5700
ctgcctttag	tgaaaaacaa	tggctagaga	gtgaagttat	gatcatgtat	agagaggtaa	5760
agttacattt	ttatattctg	actctgctaa	tgtgaaattc	cctatctgct	agactaaaag	5820
tttcagacac	cctgttcaaa	tatcccatga	ggtgctagag	acttaaaatg	aacagaacgc	5880

18/122

acattgtcag	gatgactatt	accaaaaaat	caaaagacag	caagtattgg	tgaggatgta	5940
gagaaactgg	aactttttgtg	cactgtttat	gagaatgtaa	aatggagcag	ctgctgtgga	6000
aaagagtatg	cagggttcctc	aaagagtaaa	accaagatgt	ggaaacaact	aaatgccccat	6060
cagtggtatga	aggggtagac	aatatgtggt	atatacatatc	catggagtag	tattcagcct	6120
ctaaaaaaa	aaaaggaaat	tctataacat	gcaacagcat	ggatgaatct	tgaggacatt	6180
ttgctaataga	aataaggcag	tcatagaaaag	acaaatactg	cacgactcca	cttatatgag	6240
ataccaaaaa	tagacaaatt	catagaatca	aagagtacaa	tggaggttac	ctggagctgc	6300
agggcgggaa	acgaggaggt	actaatcaac	gaacataacg	ttgcagttaa	gtaagatgaa	6360
taagctctca	agatcagctg	tacaacactg	tacctagagt	caacaataat	gtattgtaca	6420
cttaaaaaatt	tgtaaagggt	agattaacaa	atgtagtata	tccacaaatg	tggtaaagt	6480
ttcttaccac	agtaaaataa	aaaaagaata	tcaagcccag	gagttcgaga	ctagcctggg	6540
taacatggtg	aaaccctgtc	tctacagaaa	atacaaaaaat	tagccagctg	tggaggtgca	6600
ctcctagggg	ggctgagggt	ggaggccttg	ttgagcccag	gaggtcaagg	ctgcagtgag	6660
ccatgattgc	accactgtac	tccagcccag	atgacagagc	aagacaccac	cccccccaaa	6720
aaaagabaaa	gaattatgaa	catttttaaaa	gatcagatac	gcaagaacaa	caacaaaaaa	6780
gagatgaaca	gagcatcgac	cctcatctag	tgggattctt	ggtctaactg	aaaaacagac	6840
attgagagac	aaacaatgac	agtgatgtga	tcacagcaat	tacacaggta	tccccctggg	6900
actgcagag	aaaggaggaa	tgcctaactt	tcagaaaata	gagaaagcgt	caaacagttg	6960
gtgaaagcct	tccaaacta	gagagaactg	cacacaccaa	atcacagaaa	gaagaaaagc	7020
cgtgggagat	tctgggaccc	accggctatt	tttgatggct	gaacaccctg	ctgcaggaga	7080
gacaggagct	ggaaagcatg	gtgggatgaa	acctcaaaac	gctttgcctg	cattgcttaa	7140
gatgactggg	cttgattaac	tctagtcaat	ggggacaatt	caatcaaaga	agaaagatgc	7200
tcaaattcac	attttagaat	gattttttat	ggcagtatgg	ggaatagatt	aaaagagagt	7260
gaagctggag	gcaugaaact	tgtaaagagg	caactgaaac	agtctagatg	ataaataata	7320
aactgacaga	gtgactagaa	aaatcagaac	aggctgaatc	aacagatacc	tagatgaaaa	7380
taacaggact	tgatcaccag	ttgtatcttg	gagaggaagg	agttgtttcc	ttgctttccc	7440
tacgactggg	aatacggaa	gttttgccgtg	tgtattgggt	atatactggt	gtgtagccaa	7500
tcactgacaa	acatttagca	gcttaaaaac	caaaggctta	tctcccagtt	tctgtgggccc	7560
aggaatctaa	gataggctta	gctggctggg	tctggctcag	agtttctcaa	gaggttgcaa	7620
tcaagatgtc	agctgggggt	gcacatctctg	aaggctcaac	tggggccgga	gggtccactt	7680
ccaaggagtt	cactcacctg	cctgacaagg	cagtgcctgg	tggtggcagg	agatctcaat	7740
tcattgccaa	gtgagcctct	ctatagcatt	gctggaacat	cctccccatc	tggcagttgg	7800
cttctctcag	ctgagtgatg	ctgagagaga	gagcaaggag	gaagccacag	tgttcttctt	7860
actcctactc	ctaacactat	ggacactactc	ctaactctct	cacttctgcc	ttattccatt	7920
agttagaaa	ggaactaagc	tccacctctt	gaaataagaa	gtgtcaaaga	atthgtggat	7980
atatttaaaa	atcatcacac	tgtggaagtg	gatagggggt	tcaattaatg	ctgaacttga	8040
aatgcctgag	acattcaaat	gtccaacagg	caatgaacat	accatagat	ggctcatgact	8100
ttagcaagaa	tagaggaaga	tcacagaatt	aaggaggaat	tgaagggtaa	aagaagtgga	8160
gtcagattcc	ccctgaaaag	tgagccatga	aaggaaacttt	aactattgag	ttagaggtca	8220
gagtaggaaa	tttcgggtgga	attctttttt	aaagaaaagga	accatataag	catgttttga	8280
ggtagagggg	gaataaatca	gtagacaggg	agaggtaaaa	aacataaatg	ataggggata	8340
gttgacaaa	gtcttggcag	aatcccttac	ccattgactt	ggggccaaga	gagggacact	8400
tctttgtttg	agggataagg	aaaataagaa	agaatgggtg	ctatttagtg	tggtcctgtc	8460
tctagggcaa	acgcataggt	aacaaactgt	gtgtgttagg	aatatagatg	tgacctcaca	8520
ttgagattct	cacctcaaat	ccatttttgt	gttacctgta	ccttctacc	ttctcttttt	8580
gctacatgca	gactgctgtt	ttgtcttctt	ggcctgttcc	aggtttcagc	attctggcat	8640
atctgctacc	ctgttcccaa	acctctctag	agtcctgtct	ccttcttggg	atagtgtttg	8700
attggggcac	gtatctaaga	agtgatgcct	tcagttaggc	ctgagaacct	cctctatgga	8760
aatctccatc	agtgaccctg	acagacttgg	tatcttggag	atgtcactgc	tcccagcctg	8820
tggtctagga	gaatctcagc	ctgggcctct	agttagtatg	ataaggcgtt	aaggatctt	8880
tgaaccagag	tctgtcatat	tcctcaatgt	gggacagata	aaacagtggg	agtgcctggg	8940
tttctgagct	agaactctgg	tttttgggtc	agattctttg	atgtatgacc	tttcagaggt	9000
attaaaaatt	gttctaatac	aatgttcaat	acaaatgtag	ttccttttct	gttaggacct	9060
caacaaaaa	tgaccaactg	tagatgaaca	ttaaactatg	acaattcatg	gaaatgaata	9120
cagtaataacc	tgcggttccc	ccatttttagc	agtcactatg	gtgacatttg	gcacaaatgg	9180
ctattttaagg	gtgcttttgt	taaaacctac	catcttacta	ggcacatgat	attgaaacta	9240
atgaaataat	ggagaaactt	cttaaaaact	tttaatgaat	aaagtgatga	agtgataata	9300
ttttagctgc	tatttataaa	gtgactatta	cagggtcaaac	attcttctag	ggtttttttg	9360
ttgaagttgt	cacattttaat	ccttaataac	ccactatgag	tcaggtattc	ttctctcccc	9420
tttggacagt	tggggaaatg	ggggtcagag	aggttaggta	atttgcctcag	ggccacacaa	9480

19/122

cctgcatgta	gaaaaatctga	gattttgtaca	ggaacgtatc	aaactctgaa	gtccatgctt	9540
ctattttccc	atgctgcctt	tctaataaaa	ggttaactaat	gctactggat	gctgccccca	9600
aagtgaagtca	ctttcacccc	accctacttg	attttctcca	taaaactaat	cacatcctga	9660
caacttattt	attgctgata	tccccacta	gattataaac	tcaataaaaag	caagatcctt	9720
gtctgctgaa	tatcagtacc	taaaacgctg	tctagcacag	agcaagtaat	taatatttgt	9780
tgaatgaaca	aataaaaggaa	aaaaattcaa	aggaagaaaa	agccctaaaa	cagatgttta	9840
cctaaacata	catttttaaaa	gaaagcatat	aacaaattca	ggacagaatt	taaatttgat	9900
tttttaaaaga	aataaccaag	tgctagctgg	gcacagtggc	tcacacctgt	aatcctagca	9960
ctctgggagg	ccgaggcagg	cagatcactt	gaggtcaaga	gttcaagacc	agcctggcca	10020
acatgggtgaa	acctgtctct	actaaaaata	cagaaattat	ccaggcatgg	tggcaggtcc	10080
ctgtaacccc	agctactcag	gaggctgagt	caggagaatt	gcttgaaccc	aggaggcaga	10140
gggtgcagt	ggccaagatt	gcaccactgc	actccagcct	gagtaacaaa	gcaagactct	10200
gtctgaagga	gaaggaaaga	aagaaggaaa	gaaggaaaga	aggaaagaag	gaaagaagga	10260
aagaaagaaa	gaaagaaaga	aagaaagaaa	gaaagaaaga	aagaaagaaa	gaaagaaaga	10320
aagaaagaaa	aagaaagaaa	gaaagaaaga	accaagtgt	tatttgggac	ctactatgct	10380
atgtttttcc	atgcacgcta	ttttcagtaa	agcagttagc	aaacttgcaa	gatcataaca	10440
acaaatatat	gcttctataa	ctctaaaatt	gtgctttaag	aagttcctct	ttaccagctc	10500
atgtatgcat	tagttttcta	agagttacta	gtaacttttt	ccctggagaa	tatccacagc	10560
cagtttattt	aaccaaagga	ggatgcttac	taacatgaag	ttatcaaatt	tgagcctaag	10620
ttgggccagt	tcatgttaat	atactccaga	acaaaaacca	tcctactgtc	ctctgacaat	10680
tttaacctgaa	aatttcatttt	ccacattacc	aaggagccag	ggtaggagaa	tatagaaaaga	10740
ccaccaagaa	atccttactt	ctttcagcaa	aatcaattca	aagtaggtaa	ctaaacacat	10800
gccctaacaa	tgaatagcag	attgtgctca	gaagaatgat	ctacaacatc	ttactgtgaa	10860
ggactactg	aatatttcca	ataagacttc	tctccaaaat	gattttattg	aatttgcatt	10920
tttaaaaaata	ttttaagcct	aaatttttaa	aggtttgata	ttggtacatg	aatagacaaa	10980
cagacatgga	ctagaccaag	aattaggttc	aaacatatac	aggaatttaa	tatacgataa	11040
atctagtatt	ccaaaggaac	caacaaatgg	tggtcagaca	gcaggatagg	catcaggaaa	11100
aacacagttg	ggcaccctac	cttactccta	acaccaggag	taactgaagg	agcaccaaat	11160
atttatattt	tttaattata	gttttaagtt	ctaggggtacg	tgtgcacaac	atgcaggttt	11220
attacatagg	tatacatgtg	ccatgttggt	gaggagccac	aaatatttaa	aagaaaaaaa	11280
ttggccaggg	gcggtggctc	acacctgtaa	tcccagcact	ttgggaggcc	aaggtgggca	11340
gatcacctga	ggtcgggagt	tcgagaccag	cctgagcaac	atggagaaac	cccattctcta	11400
ctaaaaatca	aaaattagcc	aggcatgggt	gcacatgcct	gtaatcccag	ctacttggga	11460
ggctgaggca	ggagaatagc	tttaatctgg	gaggcaccag	ttgcggtgag	ctgagatatt	11520
gcactccagc	ctgggcaaca	agagcaaaac	ttcaactcaa	aaaaattaat	aaataaataa	11580
aaataaagaa	agaaaagaaa	aaaatgaaaa	tagtataatt	agcagaagaa	aacaccgtag	11640
aatcctcgga	ctcttaggat	ggggaaatgac	tataatataa	aaaccctgaa	gttataaaaag	11700
agaaaatcac	ctacatacaa	accaaactct	tctacatgcc	taaaacatag	cacaaacaca	11760
gctaataaat	catagctgaa	tgaactggga	aaacaaaact	tgactcatat	ccagacagag	11820
ttaattttcc	tacacataaa	gagtacctat	ataaaaccaa	caaaaaaacc	accactaacc	11880
caaaaataaaa	atgtgacagg	taatgaacag	gtagttcaca	gagaatacaa	atggctcttc	11940
ggcacataag	atgctcagac	tgacttttac	ttattttatt	tttgagagac	agggtctcac	12000
gatgttgccc	agggttaggt	caaactcctg	ggctcaaatt	atagtaccag	gactacaggt	12060
gtgccccacc	gcacctggct	cctcaaccac	ctgtattaac	aggaaatgca	aaataaaaact	12120
ttcaaatcta	ttttacctat	tagaatggca	aaaatttgaa	aaacttcaaa	catcatcatg	12180
ttggtgagaa	tgtgaggaga	ctggcactct	cattttttgc	tgatagcata	tataactga	12240
tggcttctat	ggaaaagcaat	ctggcagcgt	ctatcaaatt	tacaagtga	tatatccttt	12300
gacaaagcaa	ttccactcta	ggaatgtgtt	ctatatgggt	gtgcttcctg	gggtgggaa	12360
ctgggagcta	agggacaggg	gcagaagata	atcttctttt	ccctccttcc	ccgttaaaca	12420
tggtgaattt	tataactgt	aatatattat	ttttcacaaa	agataatttt	taagcgatat	12480
gtctgggaat	tttttttttt	cttttctgag	acagggtctc	actctgtcat	ccaggctgga	12540
atgccatggg	atgactctcag	ctgactgcag	cctcgacctc	ctgggttcaa	gcaatcctcc	12600
cacctcagcc	tcctgagtag	ctgggactac	aggcactgtc	catcatgcta	atttttgtat	12660
atacagggtc	tcactatgtt	gcccaggcta	atgtcaaact	cctaggctca	agcaatccac	12720
ccacctcagg	ctccaaagt	ctgggattac	aggcgtgagc	caccgcgcct	ggccctggga	12780
attcttataa	aagaaaaaat	atctactctc	cccttctatt	aaagtcaaaa	cagagaagga	12840
aattcaacct	ataatgaaag	tagagaaggg	ctcaacctt	gagcaacaaa	cacaaaaggct	12900
atttctgaga	caggaatttg	ctgaacaaaa	tcgagggaag	atgacaagaa	tcaagactca	12960
cttctcgggt	gggcgcagtg	gctcacacct	gtaatcccag	cactttggga	ggccgaggcg	13020
gacagatcac	gaggtcagga	gattgagacc	atactggcta	acacagtga	accaggtctc	13080

20/122

tactaaaaat	acaaaaaaatt	agccggggcgt	gggtggcaggt	gcctgtagtc	ccagctactt	13140
gggaagctga	ggcaggagaa	tggcgtgaac	ccaggaagcg	gagcttgag	tgagccgaga	13200
tcacgccact	gcactccagc	ctgggtgaca	gagcaagact	ctgtctcaaa	aaaaaaaaaaa	13260
aagactcatt	tctctagatc	ttgagccgta	ttcaaaattta	tctcagctta	gtgagagggt	13320
aaagcaagga	atatccttcc	ctgtggggccc	tgtctcttac	tgaaggaagg	taacggatga	13380
gtcaaggaca	ccaatggaga	aaagcactaa	caccattatc	tgatgaacat	tacgtgaaga	13440
agggtaagaa	gtgaagtggg	attgctgaag	aagtcagtga	aagcggacat	tcatttgggg	13500
aaatggaata	taggaaatcc	ataaaaagtga	ttaaaaagat	gttagaggct	gaggcggggg	13560
gaccacaggg	tcaggagatc	gagaccatcc	tggctaacac	ggtgaaaccc	catctctact	13620
aaaaatacaa	aaaattagcc	aggcgtgggt	cgaggcacct	gtagtcccaa	ctactcggga	13680
gactgaggca	ggagaatggc	atgaacctgg	gagacggagc	ttgcagttag	ccgagatcac	13740
gccactgcac	tccagcctgg	gtgacagagt	gagactccat	ctcaaaaaaa	aaagttagat	13800
acgagagata	aagatccaac	agacacacaa	ctgctaattc	tgaacagaa	aaaacaaatg	13860
gcacaggaaa	agaaaaattta	agatataaca	ccggaaaact	ttcctgaaat	tgagtaactg	13920
aatctatagc	ttgaaaggggt	ttagcatatg	ccaagaaaaa	tcagttagagt	ccaaccagca	13980
caagacacat	ctagcaaggc	tgggtgattct	accaacacag	agaaagaagt	gggtgaccca	14040
taatgcgga	aaaggcagac	catctgcagt	cttctccaga	acactggagt	ctgaagacaa	14100
aagaatgctg	cctactgagc	cagaaggggag	agaaagtgc	ccaacacatc	tttaccacgt	14160
tagaatgtca	cgcattatct	aaaggctgca	aaagccatga	aagacatgaa	agaacacaa	14220
catttacaac	atgaaagaac	acaagcattc	tcatactcaa	gaatccttaa	gaaaaatgta	14280
gtcctaatac	agcccactga	aagttaaatg	tacttaatgt	gctcattaat	gggaacttca	14340
tagcttcaaa	tcagtctggg	cccatctacc	aacatctctc	gcccggcttt	cctgcaatag	14400
tcagcacctt	tccttccctc	cagtcttgct	ccctggagtc	tgtctctcag	atagcagagt	14460
gaccacatca	acacccaagt	cagagccctc	cagtgcgac	tgggtctacaa	agcccttccc	14520
acccccacc	ccacgtgccc	tccggatcct	tgtgacgtgt	ctcctgcata	ccctagcagc	14580
cctggcctcc	tcactgcccc	tcctgtacat	caggaaggcg	actccttgag	tcttgggtct	14640
ggcgcctcc	tccacctgca	gtgagttaac	tcctttacct	actctagggt	attgctcaaa	14700
tgtcagcatc	tcaatggggc	cctccctgac	taccctattt	aaattctaca	tactcccctt	14760
gaccccatgg	acctcactca	ccctattcca	cttttattct	tacaatttag	cacttggtct	14820
cttctaactg	attctaagac	ttactcattt	attacattgt	ttgccacccc	ctctagtaca	14880
taaactccag	aggggcaggg	atttctgtct	atttattcat	ttctttatcc	ctaggacata	14940
gaacagggca	tagttcagag	tattcaatgt	tatcaatgaa	tgaactagca	gtagtaccag	15000
ttccagttag	gcacagaatt	aaatctaaat	agaattaaat	ctcatgggtc	gggttaacta	15060
tggatagaaa	attagatata	attttaagaa	gcctagaaa	aaaaaattaa	taatgtaaaa	15120
ataatattaa	tttgataata	ataacaaaaa	ctctgccagg	cactgtgggt	caaactctgca	15180
atcccagcta	ctcaggaggc	tgagggtgaa	ggatcacttg	agaccagagt	tcaagactca	15240
gcctaggcaa	cacggcaaga	aactgtctct	aaaaaaaatt	aaacttaaat	ttttaaaaaa	15300
gaatttctca	agcgtcacaa	aaactggaga	ttaagggtaca	ggaagtgtga	agtaaatatta	15360
ctatgctaata	gggttttttt	tttttttagaa	aggtataacc	aaaagatttc	tttctcaagt	15420
cgataaaactg	agaaagataa	gcatatcttc	caattaacag	agggggaggga	aaagccagat	15480
acaacaaaaat	aagatataaa	ttagtttcca	gttgaaaaca	agagtaggag	ttattttgca	15540
tcacctcacc	tgtgacctcc	cccagcccaa	aaaacactac	tgataaacag	ggtagaaaag	15600
catcatctca	gataaagcag	gaaaaactgc	cacagtctca	aaccacaaac	tataagcaca	15660
cacctggcca	accctgcca	gtctgggctc	agtaggagga	acgtgctgag	agctaggatg	15720
taccaactta	gacattctgt	gggatacaga	tgtccctgga	agggtcacac	catctcaaag	15780
gcacctgtaa	tgcccactga	ttacagccac	catatgtgag	agagaaaact	agggcactta	15840
gagagtataa	caagaacctt	atgtcatctg	agatgaggaa	tcctcagccc	tgcaaatata	15900
ccaactcttt	agaacaactg	gcaaaacata	aatatccaca	acttttgttt	cagtaattcc	15960
actcttagat	atcaatccaa	agtacatgag	acagcagata	cacacacaaa	atgggtattta	16020
ctgcagcatt	gtttataata	gcaaaaaaca	acaaataatc	catatgtctc	aataggatag	16080
tgggtacatg	aggggtatgta	cccatcattc	aacctcaaaa	aagagtgtata	tggatgtcca	16140
cagatggaca	taaaaagctg	tgtgttacgt	gaaaacaaac	tcaagcagca	gcaggatggg	16200
cttatgatag	tcagtatgag	ctaatttctg	gaaaaaaaaa	tctagtgtgt	gcacagaaaa	16260
catctgaaa	aacagaaaaca	aaactatcag	cagaatattg	agatgtttta	ctaagttgta	16320
tatctatact	gcttgtaatt	tttaccctaa	gcaagaatta	ctttttggaa	aaagaaaaatt	16380
caggaaataa	agcattttctt	taaacttcat	gtttaaacaa	atgggtgatgg	aataaaaagag	16440
ttcttattca	tcataaacac	acacagcaca	catgcacgca	tgtgcgtgag	cacacccttt	16500
acttgataaa	taccatgttg	aatatttttg	tctttccttt	taggttctat	cccttcactc	16560
aaaatgcggt	tataataaaa	tgtacttttc	atgtgccttc	tgccctaaacc	cactttaata	16620
taactttaca	gtcccattat	cattatagtc	tcaaagctag	actcagcctg	aaactaccct	16680

21/122

ttcatttgga	acccttatta	aaatgccaca	tacagctcct	tcaaataaaa	acaaacccta	16740
ggacctgaca	ctaggcttcc	tttgttgcta	ctcataatgg	ccaagttctg	tgcttataat	16800
acatcttctt	tcatttttatt	gctacatata	caagggtttt	atatgttttt	cttattatat	16860
cttaattcaa	aacaccatca	cgctcttttc	cagatgaaaa	taaggaaaaag	aaattgagca	16920
actgactgac	ttaaagggtca	taaaactata	tagtagcaga	gtcagcaaaa	gaagaaacac	16980
acatctccca	agtagaggct	gaaaaccagt	accattcacc	tccagggtga	gctatataca	17040
gattacaaag	tcaccttctc	taaatgttca	aactgaatcc	cataccata	ctttaccact	17100
acctcgtaag	aacagcctca	gatcttggtta	tagccttttt	tttagcatgc	tgaagccaat	17160
aaaatgcttc	ccattcagca	agagaaaaca	gttctgaaac	actgaataat	ctgcccagg	17220
cctatgaaca	tttccactgt	gagaaatgtt	ctccactgtg	tggagaagat	ccttactctt	17280
ctccacacag	gcagaacatt	agaaaaattc	ttggattcta	tgatgcacag	cttaggagtc	17340
tgtttagcac	aattttaagtc	caaatagtta	ttaaatcctc	ctctgttcca	gaaacagtgc	17400
taaatactgt	gaatataaaa	attgaaaaga	tactctcctg	gctcccaaga	aagtcagcca	17460
gatagaggag	acacaggcac	acaaatcact	gtcacatgaa	gctctacctc	cctaacttca	17520
aacgagggcc	taagtcacca	agaatacagt	agcagttgtg	actacgagta	actactataa	17580
ttcaatactt	tatcttccct	tagaaaaactc	ttctcccttg	gaaatttatt	tgcatttcta	17640
aataccattc	cttactaaaa	ggaagcagg	ctccttgggg	aaatagctga	ttctagggtg	17700
ggactatgaa	atgaaaatgg	tgagtctggg	acatcccattg	ttgcccagaa	atcaaggaac	17760
tgcccaaaaga	ttaacagagt	catgtttaa	ggacctaaaga	gtgaaccaga	aggagctcac	17820
tttgccccgc	gtggaacaat	ttcaagaaaa	acatgacagt	aatgaattat	aaaacatgaa	17880
ttaaaataca	tatttggtact	aaaaagagaa	caaaaggatg	tggctttgga	taaagctctt	17940
cttcatggaa	gaataccagc	taataaatgt	aaaggaaatg	agagaattag	aaaaattatc	18000
attttgtaaa	ccttaatatata	ttcacctaga	catgctaaaa	ccactgagta	aaaggctgct	18060
tgggaagagg	atgtctcacat	gatctcagag	tttcacacca	cagataattt	attagatata	18120
ggaagggaaga	tgtgatcaag	cttcctgtga	ccccagcca	ggccccacaa	cactatgtgc	18180
ctccttgtga	tgtgggagct	acacagcatc	gccccacacag	cttctcgcca	aaactgtttg	18240
aagctaataca	caagggaaga	actggacagc	ttctgaccat	gagacgctcc	accagacaac	18300
ttgcttgccc	tctccaaaga	aacttgcttg	gcctctccaa	agaaaactca	gtttctattta	18360
aaaacaaaac	taattatttta	aaaacaaacg	aaaagcaagt	tgtggacttg	agctccagg	18420
acagagcaga	catacttttc	cctgttcttc	ccagtaagtg	gtaataaaaa	ccctcaacac	18480
tagatataaa	acaaatataa	gaagggtctg	gaagggaag	aggaggcaga	ctatccagg	18540
gccttgaggc	ccacagaaca	acccagtgat	gggttccactg	ggtcttcttt	ttgcttcatt	18600
atctcagact	tggagctgaa	gcagcaggca	acttcaaaac	accaaggggc	acagattgaa	18660
aagccccaag	aaaagcctgc	cctctctagc	ggaaggagac	ggaaggagac	agtctaata	18720
gatggaacac	atttagacag	taactgcccc	tttaccagca	ataactgagc	agggagccta	18780
gacttccagt	cttgtgagga	cgtaccaagg	tacccaacac	ccccaccaag	gctgagtaag	18840
gactgcgagt	tttatccctg	catggcagta	gtaaggagcc	catccctcac	ccgccagcag	18900
tgtcagggga	acctggactt	ccactccac	ccaggagtg	tgaggccctc	cctgtgggg	18960
tcatgtcaga	ggaggcctag	tggagattca	gtgacttaac	cttttcccag	agataatgag	19020
gccacctttc	ctccctcttc	ccccatggtg	acagtgaag	cactgtggca	agcagtaggc	19080
actcctaccc	ctcctagcca	gggaggtatc	agggaggcca	agtagggaac	cagaataccc	19140
acaaccaccc	agcagcaaca	ggggtccccc	accccatgtg	gtgtcaatgg	aagcagagcg	19200
gaaagcctgg	atatttaccc	ccatctagaa	gtaacaagct	gatgtccccc	ttcttctact	19260
acaatgggtg	tcaaaacagg	tttaaataag	gtctagagtc	tgataacgta	atacccaaat	19320
cggtgaagtt	ttcattgagg	atcatttata	ccaagagtca	ggaagatccc	aaactgaaag	19380
agagaaaaga	caattgacag	acactagcac	taagagagca	cagatattag	aactacctga	19440
aaggatgtta	aagcacatat	cataagcctc	aacaggctgg	gcgcggtggc	tcacgcctgt	19500
aacccccagca	ctttggggagg	ccgaggcagg	tggatcacaa	gatcaggaga	tcgagaccat	19560
cctgggtaac	acggtgaaac	cccgtctcta	ctaaaaatac	aaaaaaaaaat	agcaaggcat	19620
ggtgggtggg	acctgtagtc	ccagctactc	gggagcctga	ggcaggagaa	tggcatgaac	19680
ctgggaagag	gagcagtgag	ccgagatcgc	accaccgcac	tccagcctgg	gcaacagagc	19740
aagacttcgt	ccccaaaaaa	aaaaaaaaaa	aaaaaaaagc	ctcaacaaaac	aactacaaaac	19800
gtgcttgaaa	caaatgaaaa	aaaaatcttg	gcaaagaaat	aaaagatata	tattttggcc	19860
aggtgcagtg	gctcacagcc	tgtaatccct	gcactttggg	aggctgaggc	aggcggatca	19920
cctgaggtca	ggagtttgag	accagcctga	ccaacatgga	gaaaccccg	ctctactaaa	19980
aatacaaaat	tagccagtca	tgggtggaca	tgctgtgta	cctagctact	caggaggccg	20040
aggcaggaga	atcgcttgaa	ctcaggaggt	ggaggttgcg	gtgagccgag	atcccgccat	20100
tgcacattgc	actccagcct	gggcaacaag	agcaaaactc	catctcaaaa	aaatagatac	20160
atattttaat	ggaaatttta	gaattgaaaa	atacagtaac	caaattgaat	ggaaagacaa	20220
catagaatgg	agggggcaga	caaaataatc	agtgaacttc	aacagaaaat	aatagaaatt	20280

22/122

acccaatatg	aagaacagaa	agaaaaataga	ctggccaaaa	aataaagaag	aaaaaagagg	20340
agcagcagga	ggaatgatgg	aaaaagagaa	aggaaggaag	gaaggggaagg	agggagggaa	20400
ggagtgaggg	agaaagtctc	aaagacctct	gagactaaaa	taaaagatct	aaccttgtc	20460
atcagggtcc	aggaaagaga	caaagatggc	acagctggaa	acgtattcaa	aaaataatag	20520
ctgaaaactt	cccaaatttg	gcaagagaca	taaacctata	gattcgaaat	gctgaacccc	20580
aaataaaaaag	cccaataaaa	tccacaccaa	aatacatcat	agtcaaactt	ctgaaaagac	20640
gaaaagagaa	aacgtcttga	aagcagtggag	tgaacaaca	cttcatgtat	aagggaaaaa	20700
caattcaagt	aacagatttc	ttacagaaat	taaggaagcc	agaaggaaat	gacacaatgg	20760
ttttcaagtg	ctgaaagaaa	agaagtgtca	acacaaaatt	ctagatttcag	taaaaatatc	20820
cttcaagaat	caatgggaaa	tcaagacagt	ctcagataaa	gcaaaaataag	agaatatgtt	20880
gccagcagat	ctccccataa	ggaatggcaa	aagggaagtc	atgcaacaga	ccaaaaaatg	20940
atgaaagaag	gaatccagaa	acatcaagaa	gaaagaaata	acatagtaag	caaaaaataca	21000
tgtaatcata	ataaaaatttc	tatctcctct	taagacttct	aaatttatatt	gatgggtgaa	21060
gcaaaaatta	taaccctgtc	tgaagtgtct	ctactaaatg	tatgcagaga	attataaatg	21120
gggaaagtat	aggtttctat	acctcattga	agtggtaaaa	tgacaacact	gtgaaaagtt	21180
acatacacac	acacacgtaa	gtatatataa	atatatgtgt	gtatatgtgt	gtgtatatat	21240
atatctacat	ataatgtaat	acagcaacca	ctaacaacac	tatacaaaga	gataataacc	21300
aaaaacaatt	tagataaatt	gaaatggaat	tctaaaaaat	attcaaatac	tctacaggaa	21360
gacagacaa	aaagagaaaa	aaagaggagg	acaaaactaa	tttttttaaaa	acataaataa	21420
aatggtagac	ttaagcccta	acttatcaat	aattacataa	atgtaaatga	tctaattata	21480
tcaattaaaa	gacagagata	gcagagttaa	tttaaaaaca	tagctataag	aaacctgctt	21540
tgggctgagt	gcagtgactc	acacttgtaa	tcccagcact	tcgggaggcc	aaggcggtg	21600
gatcacctga	ggtcaggagt	tccagaccag	cctggacaac	atggtaatac	cccattctta	21660
ctaaaaatac	aaaaaaatta	gccaggcatg	gtggcacacg	cctgtagtcc	caactactca	21720
ggaggctgag	acacaagaac	tgcttgaacc	cgggcagcag	aggtagcagt	gggccaagat	21780
tgcgccactc	cagcctgaac	gacagagtga	gactccacct	cagttgaaaa	acaaaaaaga	21840
aacctgcttt	aaatatacca	acatatgttg	gttgaaatta	aaagaataaa	atatatcatg	21900
aaaacattaa	tcaaaaagaa	ggagtggcta	tattaataac	ataaaataga	cttcagagaa	21960
aagaaaattt	caagagacag	gaataaaaagg	atcaagaaaa	gatcctgaaa	gaaaagcagg	22020
caaatcaatc	attctgcttg	gagattcaac	accctctctt	aacaactgat	agaacaacta	22080
gacaaaaaaa	tcagcatgga	gttgagaaga	acttaacacc	actgaacaac	aggatctaata	22140
agacatttat	ggaacactct	acccaacaat	agcaaaaata	acattctttt	caagtattca	22200
ctgaacatat	ccttagaccc	taccctgggc	cataaaaacaa	agctcactag	tgattgccga	22260
aggcttggat	ggacagtgga	agagctgcat	ggggaggggag	aagggtgacag	ttaaagagtg	22320
taggatttct	ttttgggata	atgaaaatgt	tccaaaattg	attgtgggtga	tgttggcgca	22380
actctacaaa	tataaaaaag	gccattgaat	tgtacgtttt	aagtgggtga	aacatatggt	22440
atgtggatta	tttttaaacg	tttttaaaaa	cttaacacat	ttcaaagaat	agaagtcata	22500
cagagtgtgc	tctactggaa	tcaaactaga	aagaggtaac	tggaggataa	cgagaaaagc	22560
ctccaaatac	ttgaaaactg	gacagcacat	ttctaaaatc	atccgtgggt	caaagatatt	22620
catttctgat	attcattttt	attgtttaat	gtatttttaa	aaatttctta	agggaaataa	22680
actgactaaa	aatgaatatg	gctgggtgag	gtggctcacg	cctgtgatcc	cagcactttg	22740
ggaggccgag	gctggtggat	cacaagatca	ggagtccgag	accagcctgg	ccaagatggt	22800
gaaaccccg	ctcaactaaa	aaactacaaa	aagtagccaa	gcgcagtggc	gggagcctgt	22860
ggtcccagct	acttgggagg	ctgaggtagg	agaatcgctt	gaacacaggc	agcagaggtt	22920
gcagtgaacc	aagattgtgc	cactgcacgc	cagcctgggc	gacagagact	gcctcaaaaa	22980
aaaaaaaaaaa	aaaaagaata	tcaaaatttg	tgggacatag	ttaaagcaat	gctgagaggg	23040
aaattttata	cactaaatgt	ttacattaga	aaagagaaaa	agtttcaaat	caatagtctc	23100
cactcccata	tcaagaacac	agaagatgaa	gagcaaaaata	aacccaaagc	aagcaaaaaga	23160
aagaaaatat	aaaaataaat	cagtaaaatt	gaaaacagaa	acacaataaa	gaaaatcagt	23220
gaaacaaagt	actgattctt	cgaaagatta	ataaaaattga	caaacctcta	gcaaggctaa	23280
caaacaaaaa	agaaagaaga	cacggattac	cagttattag	aatgaaagca	taattagaaa	23340
caactctaca	cattataaat	ttgacaatgt	agatgaaatg	gactaattac	tgaaaaaaca	23400
caaattacca	caactcacc	aatatgaaat	agataattgg	gatagcctga	taactactga	23460
gaaaattgaa	tttgtaat	taacactctt	aaaacagaaa	cattaaactt	aatattttat	23520
aaattattaga	taaggttaatt	atacccttcc	ttacaaaata	aaaacgacaa	attattttgc	23580
agctaaagag	atgtatgtac	tgtgaaaaat	atcttcagaa	aaatagaact	ttgtttgaag	23640
aataaggatt	taaaaaatgt	ttttaactct	caagaagcaa	atatctgggc	ccagatggtt	23700
tactgaaga	attctaccaa	atgtttaatg	aagaattacc	accaactcta	catagcatct	23760
ttgagaaaac	tgaagagaag	ggaacatctc	ccagttcatt	ttatgaagtg	ggtgttactc	23820
tgatactaga	actgtataag	gacagctact	cttgacacac	tgcctatggg	tagctctgct	23880

23/122

ctgcaggaac	agtcagaaaa	aaaaaaaaaa	gaagcactgg	acaagggcag	tataaaaaaa	23940
gaaaactggg	ccaggtgcag	tggctcacac	ctgtaatctc	agcactttgg	gaggctgacg	24000
ctggtggatc	acctgaggtc	aggagtttga	gactagcctg	gccaacatgg	taaaaccctg	24060
tctctactaa	aatacaaaaa	ttagccaggc	aggggtgggg	ggaaaaataa	aaggaaaaaa	24120
aaacaaaaat	aaactgcaga	ccaatatcct	tcatgagtat	agacacaaaa	ctccttaaac	24180
tctttaacaa	aattattagca	agtagaagca	atatataaaa	ataattatac	accatgatca	24240
agtgggactt	attccagaaa	cgcaagtctg	gttcaacatt	tgaaaaacaag	gtaaccctact	24300
atatgaacgt	actaaagagg	aaaactacat	aatcacatca	atcaatgcag	aaaaaagcat	24360
ttgccaaaat	ccaatatcca	ttcatgatag	tctaataaga	aaaataagaa	taaaggggaa	24420
attccttgac	ttgataaaagc	ttacaaaaga	ctacaaaagc	ttacagctaa	cctataactta	24480
atggtgaaaa	actaaatgct	ttcccctacg	atcaggaaca	aagcaaggat	gttcactctc	24540
attgctctta	tttaacatag	ccctgaagtt	ctaacttggt	caaaacgata	agaaagggaa	24600
atgaaagacc	tgcagattgg	caaagaagaa	ataaaaactgt	tcctgtttgc	agatgacatg	24660
attgtctcat	agaaaatgta	aagcaactag	gggtaggggg	gcagtgaggag	cacgctgggtc	24720
aaaggatacc	aaatttcagt	taggaggagt	aagttcaaga	tacctattgc	acaacatggt	24780
aactatactt	aatatatgtt	attccttgaaa	atactaaaag	agtgggtggt	aagcgttctc	24840
accacaaaaa	tgataactat	gtgaagtaat	gcatacgtaa	attagcacaa	cgtatattac	24900
tccaaaacat	catgttgtag	atgataaata	cacacaattt	tatctgtcag	tttaaaaaaa	24960
catgattttg	gccaggcaca	gtggctcata	cctgtaatcc	cagcatttta	ggaggctgag	25020
gcgagcagaa	aacttgaggt	cgggagtttg	agaccagaat	ggtcaacata	gtgaaatccc	25080
gtctccacta	ataatacaaaa	aattagcagg	atgtggtggc	gtgcacctgt	agaccagct	25140
acttggggagg	ctgaggcacg	agaattgctt	gaacaaggga	ggcagagggt	gcagtgagct	25200
gggtgccact	gcattccagc	ctggtgacag	agtgcagctc	catctcaaaa	aaaataaaaat	25260
aaagcatgac	ttttcttaaa	tgcaaagcag	ccaagcgcag	tggctcatgc	ctgtaatccc	25320
accactttgg	taagccgagg	caggcagatc	acaaggtcag	gagtttgaga	ccagcctgac	25380
caacatgggtg	aaaccccatc	tctactaaaa	aatatataaa	ttagccaggc	atgtgtagtc	25440
tcagctactc	aggaggctga	ggcaggagaa	tcacttgaac	ccggaggcag	agggtgcagt	25500
gttgagccac	cgcactccag	cctgggtgag	agaacgagac	tccgtctcae	aaaaaaaaaag	25560
caaaataaacc	taatttttaa	aacactaaaa	ctactaaagt	aattcagtaa	gtcttttagga	25620
ttcaggatat	atgatgaaca	tacaaaaaatc	aattgagctg	gacaaaggag	gattgtttta	25680
ggtcagtagt	ttgaggctgt	aatgcacaat	gattgtgcct	gtgaatagct	gctgtgctcc	25740
agcctgagca	gcataatgag	accacatctc	tatttataaaa	aaaaaaaaaatt	gtatctctat	25800
gtactagcaa	tactactaaa	ggtactaaaa	ttaaaaacat	aataaatact	gttttttaatt	25860
gcctgaaaaa	aatgaaatag	ttacatatata	atctacaaaa	atgtgcagga	cttggtgtgct	25920
gaaaactaca	aaacgctgat	aaaagaaatc	aaagaagact	taaatagcgt	gaaatataacc	25980
atgcttatag	gttggaatac	ttaatatagt	aaagatgcc	attttatcca	aattattaca	26040
caggataaca	ttattactac	caaaaatccc	gaaaaatctt	acatagatat	agacaagatc	26100
atacaaaaaa	gtatagcgaa	atatgcaaa	gaactagagt	agctaaaaaa	aatttgaaaa	26160
agaaaaataa	agtgggaaga	atcagctctat	ccagtttcaa	gacttacata	gctacagtaa	26220
tcaagactgt	gatattgaca	gagggacagc	tatagatcaa	tgcaaccaaa	tagagaacta	26280
agaaagaagc	acacacaaat	atgccccaat	gattttctgac	aaaggtgtta	aaacacttca	26340
acggggggaag	atatgtctct	cattaaaggg	tgtagagtca	ttgcacatct	ataggcaaaa	26400
agatgaacct	gaacctcaca	ccctacagaa	aaattaaactc	aaaatgactc	aaggactaaa	26460
cataagatat	acatctataa	aacatttaga	aaaaggccac	gcacggtggc	tcacgctcgt	26520
aatcccagca	ctttggggagg	ccaaggcagg	tggatcacct	aagggtcagga	gtttgagacc	26580
agccggatca	acatggagaa	gccccatctc	tactaaaaat	acaaaattag	ctggacgtgg	26640
tggcacatgc	ctgtaatccc	agctacttgg	gaggctgagg	catgagaatc	gcttgaacct	26700
gggggggcaga	ggttgcggtg	agccaagatc	acaccattgc	actccagcct	gggcaacaag	26760
agcaaaaactc	caactcaaaa	aaaaaaaaaaa	aaaggaaaaa	tagaaaatct	ttgggatgta	26820
aggcgaggta	agaatttctt	acacttgatg	ccaaactaag	atctataagg	ccagtcgtgg	26880
tggctcatgc	ctgtaattcc	agcactttgg	tcaactagat	gaaaggtata	tgggaattca	26940
ctgtattatt	ctttcaactt	ttctgtaggt	ttgacatttt	tttagtataaa	aattggggga	27000
aagacctgac	gcagtggtctc	acacctgtaa	tcccagcact	ttgggaggcc	ggggcaggtg	27060
gatcacacgg	tcaggagttc	gagaccagcc	tggccaacat	ggtgaaaccc	cgtctctacc	27120
aaaaatataa	aaaatttagcc	gggtgtcatg	gtgcatgcct	gtaatcccag	ctactgagga	27180
ggctgaggga	ggagaatcac	ttgaacctgg	ttgagtggaag	ttgcagtgag	ccgagattgt	27240
gccactgcac	tccagccttg	ggtgacagag	cgagactccg	tctcaaaaaga	aaaaaaaaaaa	27300
aaagaatatc	aaacgcttac	tttagaaact	atttaaagga	gccagaattt	aattgtatta	27360
gtatttagag	caatttttat	gctccatggc	attgttaaatt	agagcaacca	gctaacaatt	27420
agtggagttc	aacagctggt	aaatttgcta	actgttttagg	aagagagccc	tatcaatatac	27480

24/122

actgtcattt	gaggtcgaca	ataagcacac	ccaaagctgt	acctccttga	ggagcaacat	27540
aaggggttta	accctgttag	ggtgttaaat	gtttggatat	ggtttggttg	gccccaccga	27600
gtctcatgtt	gaaatttggt	ccccagtact	ggaggtgggg	ccttatttga	aggtgtctga	27660
gtcatggggg	tggcatatcc	ctcctgaatg	gtttggtgcc	attcttgcag	gaatgagtga	27720
gttcttactc	ttagttccca	caacaactgg	ttattaaaaa	cagcctggca	ctttccccc	27780
tctctcgctt	cctctctcac	catgtgatct	cactgggttcc	ccttcccttt	atgcaatgag	27840
tggaaagcag	ctgaagccct	cgccagaagc	agatagtgat	gccatgcttc	ttgtacagcc	27900
tacaaaaacca	tgagcccaat	aaaccttttt	tctttataaa	ttatccagcc	tcaggtattc	27960
ctttatagca	agacaaatga	accaagacag	ggggaaatca	acttcattaa	aataatctat	28020
gcagtcacta	aacaaataag	aacaagaggg	tccagaagtg	ggaagccaat	acccagagtt	28080
cctacaatac	agtatctgaa	aagtccagtt	tccaaccaaa	aaatatatat	atacaggccg	28140
gacatggtag	cttatgtctg	taatccagc	actttgggat	gctgaggcgg	gcagatcacc	28200
ctaggtcagg	agttcgagac	cagcctggcc	aatatggcaa	aaccccgctc	ctactaaaaa	28260
tacaaaaatt	agccaggcat	ggtggtggat	gcctgtaatc	ccagctactc	gggaggctga	28320
ggcagggaa	cacttgaacc	caggaggcag	aggttgcagt	gagccgagat	cacgccactg	28380
aactccagcc	tgggcaacaa	agtggagctc	cacctcaaaa	aaaaaaaaaa	tatacatata	28440
tatatgtgtg	tgtgtgtgtg	tgcgcgcgtg	tgtgtatata	cacatacaca	tatacatata	28500
tatacagaca	cacatatata	tatgaagcat	gaaaagaaac	aaggaagtat	gaaccatact	28560
ttctgtgggt	atgataggat	gggggtatcac	gggggaagta	gacaagggaa	actgcaagtg	28620
agagcaaaca	gttatcagat	ttaacagaaa	aagactttgg	agtaaccatt	ataaatattg	28680
ccacagaatt	aaagaaaagc	gtgattaaaa	aaggaaaagga	aagtatcata	acaatattac	28740
tccaaataga	gaatatcaat	aaaggcatag	aaattataaa	atataatata	atggaaattc	28800
cggagttgaa	aggtagaata	actaaaattt	aaaattcact	agagaagggt	caacactata	28860
tttgaactgg	cagaagaaaa	atttagtgag	acaaatatac	ttcaatagac	attattcaaa	28920
tgaaaaataa	aaagaaaaaa	gaatgaagaa	aaataaacag	aatctcagca	aaatgtggca	28980
caccattaat	cacattaaca	tatgcatact	gagagtaccg	gaagcagatg	agaaagagga	29040
agaaaaaata	ttcaaagtat	ggccagtaac	ttcctagatt	tttgttttaa	agcaataacc	29100
tatacaatca	agaaactcaa	tgaattccaa	gtaggataaa	tacaaaaaga	accacaaaca	29160
gatacaccat	ggtaaaaatg	ctgtaagtca	aaaacagaga	aaatattgaa	agcagctaga	29220
ggaaaaactta	taagagaacc	tcacttacia	aagaacatca	cttataaaaag	aaccacaata	29280
atagaaacag	ttgacctctc	atcagaaaca	atgaatgata	acataatttga	agtgtctcaa	29340
gaaaaaaaat	aaagattcct	atatacgaca	aagctgtctt	tcaaaaatat	acatccaaaa	29400
ggattgaaac	cagggtcttg	aagagttatt	tgtacatcca	tgttcatagc	agcattattc	29460
acaatagcca	aaaggtagaa	gcaacccaag	ggtccactga	caaataaata	aaatgtggta	29520
tatgtatata	caatggaatt	tattcagtat	taaaaaggaa	tgaattcttg	acacatgcta	29580
caacatggct	aaaccttgag	aacactatgc	taagtgaat	aagccagcca	caaaaggaca	29640
ataaccatat	tacttcaact	gtatgaaata	cctagggtag	tcaaattcag	agatagaaag	29700
taaaacagtg	gttgccaagg	gctgagggag	ggagtaacct	ggagttattg	ttgaatgggt	29760
acagaatttc	agttttgcaa	gataaaaaaga	gttctggaga	cagatgggtg	tgaggggtgt	29820
acaacaatac	aaatatactt	tatactactg	aacagtatac	ttaaaaatga	ttaacatggt	29880
gaaaccccg	ctctactaaa	aatacaaaaa	aattagctgg	gtgtgggtgg	gggcacctgt	29940
aatcccagct	acttggggag	ctgagggcag	agaattgctt	gaaaccagaa	ggcggagggt	30000
gcagtggagc	gagattgcgc	caccgcactc	tagcctgggc	aataagagca	aaactccgtc	30060
tcaaaaaata	aaaaataaaa	aaaattttaa	aatgattaag	caggaggcca	ggcacgggtg	30120
ctcacaccta	taatgccagc	actttggggg	gccgaggcag	gcgatcactt	gagaccagga	30180
gtttgagacc	agcctggcca	acatggcaaa	accctgtctc	tgctaaaaat	acaaaaatta	30240
gccaggcatg	gtggcatata	cttataatcc	cagctactgg	tgagactgag	acacgagaat	30300
tgcttgaacc	caggaggcag	agattgcagt	gagtcgagat	cgcgccactg	aattccagcc	30360
tgggcgacag	agcaagattc	tgtctcgaaa	aaacaaaaac	aaaaacaaaa	agcaaaacca	30420
aaaaataatt	aagcaggaaa	cgagattgct	gctgaggagg	agaaagatgt	gcaggaccaa	30480
ggctcatgag	agcacaaaaa	ttttcaaaaa	atggtttaatg	attaaaaatg	taaaatttat	30540
atgtatctta	ccacaaaaaa	aagggctggg	gggcaggaaa	tgaagggtgaa	ataaagacat	30600
cccagagaaa	caaaagtaga	gaatttggtg	ccttagaaga	aacaccacag	gaagttcttc	30660
aggctgaaaa	caagtgaccc	cagagggtaa	tctgaattct	cacagaaaat	tgaagcatag	30720
cagtaaaagt	tattctgtaa	ctatgacact	aacaatgcac	attttttcct	ttcttctctg	30780
aatgatatta	aaaagcaatt	gcataaaaa	ttatatataa	agcctattgt	tgaacctata	30840
acatatatag	aaatatactt	gtaatatatt	tgcaataaac	tgcaaaaaag	agagttggaa	30900
caaagctgtt	actaggctaa	agaaattact	acagatagta	aagtaatata	acagggaact	30960
taaaaataaa	atttttaaaa	attttaaaaa	aataattaca	acaataatat	ggttgggttt	31020
gtaatatata	tagacataat	acaaaaatac	cacaaaaaag	gaagaagaca	atagaactac	31080

25/122

ataggaataa	catttttggt	tctaactaga	attaaattat	aaatatgaag	tatatctctg	31140
taagttaaga	cacacatggt	aaaccctaga	tactaaaaag	taactcacat	aaatacagta	31200
aaaaaataaa	taaaataatt	aaaatgtttg	tatttagtttc	ctcagggtac	agtaacaaac	31260
taccacaaat	tgagtggcct	aacacaactt	aaatgtattt	tctcccagtt	ctggaggcta	31320
aacacctgca	atcaagggtga	gtacagggcc	atgctccctg	tgaaggctct	aggaaagaat	31380
cctcccttgt	ctcttccagc	ttccagtggg	tctcagtaac	cctaagtgtc	ccttggcttg	31440
tagctatata	attcctagca	accagaaaga	agaaaaataa	aaagattatg	gcaaaaaata	31500
atgaaatcaa	aaggagaaaa	atggaaaaaa	ataaataaaa	ccaaaagcta	gttctttgaa	31560
aagatcaacc	aagttaacaa	accttttaac	tagactgaca	aaaaggaggt	aagactcaaa	31620
ttactagaat	cagaaataaa	agaggggaca	ttactaatga	gggattagaa	aagaatacta	31680
cgaacaaatg	tgtgccaaca	aattagaaaa	cttagatgaa	atggacaggt	tcctaggaca	31740
acatcaacta	ccaaaattta	ctcaagaaga	aagagacaat	ttgaatgagc	tataacaagg	31800
gaagagactg	aattgacaac	caagaaacta	tccacaaaaga	aaatcccagg	cccagaagat	31860
ttactgtga	aacttttca	aacttataaa	tataaattaa	catcagttct	tcacaaactc	31920
ctccaaaaaa	aagaacagat	ctctattttac	aggcgatacg	atccttagaa	aatcctaagg	31980
gaactactaa	gacactatga	taactgataa	acaagttcag	caaggctgca	ggatagaaaa	32040
ccaatatata	aaaatctatt	atatttctat	acacttgtag	tgaacaaccc	aaaaatgaga	32100
ttaagaaaaa	aattcaattt	acaataacat	caaaaagaat	aaaaacactc	aaaaataaat	32160
ttattcaagt	aagtgcacaaa	cttatactct	agaagctaca	aaacactggt	aaaagaaatt	32220
aaaggtttac	ataaatgaaa	aactatccca	tgttcatgga	tcaaaagact	tattactggc	32280
aatgctctcc	aaattgatct	ataaattcaa	caaaatcctt	atcaaaatcc	cagatgaggg	32340
tgggggtggc	ggttcatgcc	tgtaatccca	gcactttggg	aggctgaggg	acgcagatta	32400
cctgaggtcg	ggagctcgag	atcagcctga	ccaacatgga	gaaaccctat	ctcttctaaa	32460
aatacaaaa	tagtcaggcg	tgggtggcaca	tgccataaat	cccagctact	cggggaagctg	32520
aggcaggaga	atcgcttgaa	cccaggaggc	agaggttgca	gtgagccaag	atcggtccat	32580
tgcactccag	cctgggcaac	aagagcaaaa	ttccatctca	aaaaaaaaaa	aaaaaaaaatc	32640
ccagatgact	tcactgttga	aattgaaaaa	attattctaa	aattcacatg	gaattgcaag	32700
accttgagaa	tagccaaaaa	aaacttgaaa	aacacgaaca	aaatatagga	tgactcactt	32760
gccaattgca	aatgttacga	cacagcaaca	gtaatcaaga	ctgtgtggta	ctggcaaaag	32820
acacatacat	acatacatat	caatggaata	taattgagag	tacagaaaca	agcctaaaca	32880
tctatggtaa	gtgcttttct	atTTTTTTCT	TTTTTTTTTT	ctTTTTTgta	gagatagaat	32940
ctcaccatgt	tgcccaggct	ggtcttcaac	ttctgggctc	aagcaatcct	cccactgtgg	33000
cctcccaaa	tgctgggata	actggcatga	gccaccacat	ccagcccaga	tgattttcaa	33060
aaaagtcaac	aagaccattc	ttttcaacaa	ataggtctgg	gatgatcaga	tagtcacatg	33120
aaaaaaaaaa	tgaagttgga	ccctccatca	cactaaagtg	ctgcgattat	aggcatcagc	33180
caccacatcc	agcccaaatg	atTTTTCAAAA	aggtcaacaa	gaccattctt	ttcaacaaat	33240
aggctctggga	taactcagata	gtcacatgaa	aaaaaaaatg	aagttggacc	ctccatcaca	33300
ccatatgcaa	aaattaattc	aaaaatgaat	tgatgactta	aacgtaagag	ttacgactgt	33360
aaaactctta	gaaggaaaca	tacgggtaaa	tcttaaagac	gttaggtttg	acaaagaatt	33420
cttagacatg	acacaaaaag	catgaccaac	taaggtaaaa	tagggtaaat	tgtacctacc	33480
aaaatgaaaa	acctttgtgc	tggaaggagc	accatcaaga	aatggaaagc	caaaatagcc	33540
aaggcaatat	taagcaaaaa	gaacaaagct	ggaggcatca	tactacctga	cttcaaaaga	33600
acagtaacca	aaacagcatg	gtactagtag	aaaaacagac	acatagacca	atggaacaga	33660
ataaagaacc	caaaaataaa	tccacatatt	tatagtcaac	tgatttttga	caatgacacc	33720
ccttcaataa	atgatactag	gaaaactgga	tatcgatatg	cagaagaata	aaactagacc	33780
cctatctctc	accatataga	aaaatcaact	cagactgaat	taaagacttg	aatgtaagac	33840
ccaaaactat	aaaactactg	gtagaaaaaca	taaggaaaaa	cgcttcagga	cattgggtcca	33900
ggcaaagatc	ttatggctaa	aacctcaaaa	acacaggcaa	caaaaacaaa	aatggaaaaa	33960
tagcacttta	ttaaactaaa	aagctcctgc	acagcaaagg	aaacaacaga	atgaaaagac	34020
aacctgtaga	atgggagaaa	atatttgcaa	actatccatc	catcaaggga	ctagtatcca	34080
gaacacacaa	gtgactaaaa	caactcaaca	gcaaaaaagc	aaataatctg	gtttttatat	34140
gggcaaaaaga	tctgaataaa	cattctcaaa	ggaagacata	caaatgtcac	tatcattctg	34200
ccagtaccac	actgtcttga	ttacttgtaa	gtgtataaat	ttttaaattg	ggaagtgtga	34260
gtcatcctac	actttgttct	tgTTTTTcaa	gtttgttttg	gctatttctgg	gagccttgca	34320
agtataaaa	agccaacaa	tatgaaaaaa	tgctcaccat	cactaatcat	cagagaaata	34380
aaaatcaaga	ccactatgag	atatacctctc	actccagtta	gaatggctac	tatcaaaaag	34440
acaaaatata	atggatgctg	gcaaagattt	ggagaaaagg	gaactcctat	acactgtggg	34500
tagggatgca	aattggtaat	ggccattatg	gaaaaataa	ctgaggtttt	tcaaaaaact	34560
gaaaatagaa	ctaccatatg	atccagcaac	cctactactg	ggtattttatc	caaaggaaag	34620
aagtcagtat	actgaagaaa	tatatgcact	ctcatgttaa	ttgcaacact	gttcacaaca	34680

26/122

gccaagacag	ggaataaatc	taaatgtgca	tcaacagatg	aatggataaa	gaaaatgtgg	34740
catatacact	caatagaata	ctattcagcc	attaaagaag	aatgaaatcc	tgtcatccca	34800
gcaacatgga	tgaacctgga	ggacattata	tttaatgaaa	taagttaaagc	acaaaaagat	34860
aaacagtaca	tgttctcact	cagacatggg	tgctaaaaag	aaaatggggg	cacagaatta	34920
gaaggggagg	cttgggaaaa	gttaatggat	aaaaatttac	agctatgtaa	gaagaataag	34980
ttttagtgtt	ctatagaact	gtagggcgag	tatagttacc	aataacttat	tgtacatggt	35040
caaaaagcta	gaagagattt	tggatgttcc	cagcacaaag	gaatgataaa	tgtttgtgat	35100
gatggatatc	ctaattaccc	tgattcaatc	attacacatt	gcatacatgt	atcaaattat	35160
cactctgtac	ctcataaata	tgtataatta	ttacgtcaac	aaaaaaagga	aaaaaaagaa	35220
aattaagaca	acccacataa	tggagaagaa	aaaatatctg	caaattatat	atatctgata	35280
aatattttaat	at ttataata	tataaagaac	tcctacaact	caagaacaac	aacaaaacaa	35340
cccaattcaa	aaatgggtaa	aagccttgaa	tatacactta	tctaaagact	atatacaatt	35400
ggccaataaa	gacacgaaaa	gatgctcaac	atcactagtc	atcagggaaa	tataaatcaa	35460
aaccacaatg	tagaattgtag	acaccacttc	atatgcacta	ggatggctag	aataaaaaagg	35520
taataacaaa	tgttggttaag	gatgtgaaaa	aatcagaaac	ctcattcgct	gctgttggga	35580
atgtaaagtg	atgcagccac	tttggaaaac	agtctggcag	ctcctcaaat	tattaaatac	35640
agagttaccg	tatgacccag	gaatattcct	cctgggtcta	taacccaaaa	aatgaaaaca	35700
tatatccaca	taaaaacttg	tacatgggca	tttatagcaa	cattattcat	aacagcaaag	35760
gtggttaagaa	cccataatgcc	catcatctga	tgaacaggtg	aataacatgc	ggtattattcc	35820
atacactaga	atattatctg	cccatacaag	gagtgcacac	cagctacatg	ctacaaggat	35880
gaatctcgga	aaccttatgc	taagtgaag	aagccagtca	caaataacca	cagattatga	35940
ttccatgcat	cggaaatgac	cagaataggg	aaatctatag	agacagaaag	tagattagtg	36000
gttgggtggg	gctggggagga	caggtagtac	actactttcc	cagaactact	ggaacaaagt	36060
accacaaaact	ggggagctta	aacatagaaa	ttgatttccct	cacagttctg	gagactagga	36120
ctctgagatc	aagggtgtcag	cagagctggg	tctttctgag	ggccctgagg	caaggctctg	36180
tcccaggcct	ctctccttgg	ctggcagggtg	gccatcttct	ccctgcgtct	tcacatcatc	36240
ttttctctgt	gtgtgcccac	gtccaaaattt	tgattggctc	attctgggtc	atggccaatt	36300
gctatgcaca	aagtgaagtc	tacttccaaa	agaagggaag	agggaaacac	gactaggcta	36360
aactttatagt	catttttaatg	tccgcttttc	ctatgagatt	gtgaacacac	agaagtggg	36420
tttttatcta	cattgtgcaa	agtttaataa	gaaaaatag	attcaagaga	agcagttcaa	36480
tagcaggaat	ttaatatggg	aactaattac	aagggtttagg	gcaggactaa	aaagccagtt	36540
gggatgggtga	gccaaccacg	agattagcaa	cagtgggacc	ccatctacct	accacccatg	36600
aagctggaag	gataaaggag	gggctattat	cagagttccac	aagccagtgt	cagagtcctt	36660
ggctggagct	gggaccaccc	tagagacact	gtgcaaagca	gaaaacaagg	gggaaaaacc	36720
ctgacttctc	ccttcctccc	accttttcaat	ctcccactag	tgcttcctac	tagccatact	36780
tggccagaga	cagtgcacaag	gaacactgca	aaatgaagtt	tgtaggaatc	atctccctct	36840
gagacagaga	aatatggaag	ggtagaaaa	gaatcagagg	ataaagagaa	aaaaccctga	36900
gtactatctt	at ttatcttt	gtatctccag	tgcttaattc	gtctctcaaa	aaaggaaagc	36960
aattgagaga	aactgaaaac	tccaattgaa	atgaaagaat	ggagaattac	tggactagaa	37020
gagaagagaa	aaat ttatc	cgcataagag	aaacaagaat	ggattcacia	aggacgtgat	37080
gaatgaaaag	ctataatcag	caaagatttg	ccagagaaat	taaaaagtgg	taaaactcagc	37140
cacgctgtac	aacctgaagg	cacaatgcac	gaaaacggtt	caagaaatga	caagatttga	37200
agtcaaattc	taagtgtctt	tccagaatct	ctcaagacga	ttatatagct	accccat ttt	37260
attaaataaa	atggaaactt	actaaacttt	ccccttggtat	taaaactaaca	tatgtcctaa	37320
tagcaaacga	ttctgggaatt	cctagagtaa	aatatatttc	gtcaaaagtgt	attgtctctt	37380
taatattctg	ctgacctcct	tttgctattt	aggatatttg	tatacacatc	acacgtaaat	37440
ttgggtctata	gtttacatct	acgggcttat	actgttcttt	ttttcatttt	tttaaaattt	37500
ccaacccccca	gtatccatat	actgctctct	atcaggggta	ttttaacttt	gtaaaatcag	37560
ctgagatgct	ttccatgttt	ttttttttta	ttttctgcca	catttgaata	gcataggagt	37620
taccaccatc	aaccttggat	tatttaagca	ttcacgattc	cacgtgtgga	ttttttattc	37680
agagtctttc	ttgtcatttc	tgctatcagc	acagaaaccca	atctcagctt	tccagctata	37740
ctctcacccc	atggaaattg	cagatgaagt	tcaaaaggac	ccttgcat tta	tcctgcctcg	37800
ccctcttccc	ccttcattta	gacatcacct	tcttctagaa	cgtcttacct	gacatgccct	37860
gctcccaacc	cctgctgccc	aattgtgtgc	tctcccgtgt	cctggcctgc	catcctcttt	37920
agtaattggc	tcttccctca	tctgtctccc	caccagaca	ttaagctgaa	tagactggat	37980
ttgtgtcttg	tccatcacta	taatctcagc	acctagttac	tagtaggtac	ttaccatgta	38040
ttcattagca	aaatgttatg	tataaccttg	caccttaaaa	acaagagaag	gaagacaaaa	38100
ttaagtctta	agactatggg	ttagaacatg	gatcagaaac	tacagtctgc	agcccaaadc	38160
cagaccaaact	gaagagacca	tgttcattta	catacaacct	atagcagctt	tcacactaca	38220
ggagcagagc	taagtagttc	caagggaaca	cacggccctg	caaagccctaa	aatattttact	38280

27/122

ctatagctct	tcacagaaaa	agttttcaga	tccctcgttt	agaactcttg	ttcatatgca	38340
atttcactaa	accatagttt	tttgggtttg	tttgggtttt	tttggcaaaa	aggaatgagc	38400
cgatccagaa	aagggtgaaa	agaatgaatc	attactgctg	aaagaatgtg	cacacagctc	38460
gtcagttatt	tgctgccatg	ctgacaccca	tccaatagtg	tcatgagatg	cagcagctac	38520
tactgtgttc	tcaatgccga	gtccacccac	tccataacca	tgtccaagca	atcttgggaa	38580
catcatcacc	atgcttggtt	atccttaagg	tattgectca	catacagcag	tggctgggtca	38640
taaagtcaaa	tgacactagt	ggccaggagg	tcaagagaat	gagtgaggac	aggtgggtag	38700
gcagcccagg	ccctagcaac	agcaggagct	caccctcag	tactctagc	caggactgaa	38760
atacttttca	ccctttcaag	agagactagg	aatctggatt	tttatgtgaa	atatcttgat	38820
tactaaatgt	tgtcaacaga	catgtcaaaa	ggtaaaaacta	agtaagttca	tggggcagat	38880
tgactattca	ggttatagaa	tttaaggattc	ttatccaaca	cagataccaa	ccaaaaagct	38940
gacgtataac	atattaggag	aaactatgtg	cactgtcgaa	acatcaacaa	ggggctaattg	39000
tctaaaatag	tctatattgg	attccagttg	aaacatgggg	aaaggacatg	aacaggcaac	39060
tttgtgcaat	ggaaactcaa	aaagataaca	agcatatata	aaagcattct	caaattcagt	39120
agttaacaga	cagatgcaaa	taaaaagagg	gaaactgctg	ccgggcacag	tggctcacac	39180
ctgtatccc	agcactttgg	gaggccgagg	cgggcggatc	atgaagtcat	gagatcgaga	39240
ccatcctggc	taacatgggtg	aaaccccgtc	tctactgaaa	acacaaaaaa	ttagccaggc	39300
gtagtgtgg	gcaccagtag	tcccagctac	tcaggaggtt	gaggcaggag	aatggcatga	39360
acccagtagg	cggagattgc	agtgagccga	gacctgccca	ctgcaactca	gcctgggcga	39420
ctgagttaaa	ctccatctca	aaaaatataa	taataattat	aattataata	ataataaata	39480
gttaattaat	aaaaagagag	agactgctaa	agtctagaaa	gttgaatgat	gccaaagcga	39540
tgcaaatatc	agggccttgg	gatggccggg	tgcaagtggc	cacgcctgta	atcccaccac	39600
tttggjgggc	caaggcgggc	ggatcatgag	gtcaagagat	caagaccatc	ctggcccgaca	39660
cagtgaaccc	cggctctctac	taaaagtaca	aaaaaatata	tatatatata	tataattatta	39720
tatttatatat	atatatatca	gagccttggg	aatccttgtg	tgctgctggg	gaaggtagtg	39780
gtgcantcac	ccttgacagc	aatctggcag	tacttggtta	tattaagtat	aggcacacac	39840
cacgaccagg	cagtcctact	cctgggtcta	aatcccaaag	aattctcaca	caagtccata	39900
aggagacatg	atcagagctc	attcagcatt	actgggagtg	ggaatcaacc	tgggtgtcca	39960
tctacaggag	acgagatgga	caaaatgtgg	tgatatttaa	gaccagaatc	accaagtaac	40020
agagatgggt	ggtagtgac	aatcctaaga	tacagaataa	aggctagaac	atgatgccat	40080
tcatgttaat	taaaaataga	tgacacacaa	gcagtatacg	cgtgaccctt	gaatagcaca	40140
ggtttgaact	gcctgtgtcc	actttacatgt	ggattttctt	ccacttctgc	tacccccaa	40200
acagcaugac	caacccctcg	tcttctctct	ccccctcagc	ctactcaaca	tgaagtagac	40260
aaggatgaag	actttttatga	taatccaatt	ccaaggaact	aatgaaaagt	atattttctc	40320
ttcctttatga	ttttctttat	ctctagctta	cattattcta	agaatatggt	acataatata	40380
catcacacgc	aaaataaatg	ttaattgact	gttttatatta	tgggttaaggc	ttccactcaa	40440
cagtaggctg	tcagtagtta	agttttggga	gtcmetaagtt	atacacagat	tttcaactgt	40500
gcaggcaatc	agttccccctg	acccccctcat	tgttcacggg	tcaactgtat	atacacaaaa	40560
gtatttatatg	aacctcatta	gaatagctgt	ctatagggag	aagagaatga	gagtgggata	40620
aaacgggaatg	aacaaataaa	ccaacaaatg	cattaacaag	caaaacaaca	gaggggcttg	40680
catgggcccag	tgatgataaa	gggctaagaa	tgagaatata	attaattcaa	ttcctcacac	40740
ctgaggtcta	aaaccaagga	aagggagggc	caggcgtgga	ggctcacgcc	tgtaatccca	40800
gcactttggg	aggctgaggg	gggaggatca	caagattagg	agtttgagat	cagcctggcc	40860
aacacagtg	aagcccattct	ctacaaaaaa	tacaagaatt	acccaggtgt	ggtggcacat	40920
gcctgtagtt	agctactctg	gaggctgagg	caggagaatc	acttgaaccc	aggaggcgga	40980
ggttgacagg	agccgagatc	acaccattgc	actccagcct	gggtgacaga	gtaagactct	41040
gtctcaaaaa	aataaaaaaa	ataaaaaaac	agagaaaagg	aggaaaactag	atccaggctg	41100
actagataca	gccttttagag	ttagaaaaga	tgatttgaca	atctaagccc	acactcagat	41160
tgaatgaaat	tgaaaagcct	ttcaaactaa	aacattttaat	tacaccatct	gctgcagaca	41220
gaactcagac	aactcaaaca	ggtaatgtca	gcgtgggtgt	ttatatcacc	accttcaaca	41280
cagaataaaa	atcagctgca	tgtgaagcag	tgactagaat	gaagaaaagg	ctgcttctta	41340
cttccctcta	gtgggtctctt	ccgaaaacat	taataggcac	cagctctatg	catgtcaccc	41400
tgcaggggaga	catgggggtat	ataactatga	cttactgttc	attcctcaag	gaattcccaa	41460
tcttgtggaa	gattatacac	aatgaggcaa	caaaaactat	ccaataaaac	cacggaaaag	41520
aaggaagaa	caaagaagcc	agtgatgaaa	ggccctgtga	gcagagctga	tggccatttg	41580
gggaagaaag	accaactatg	atgggggtga	tcagggtggc	tccgtgggaa	agctgggaaga	41640
gaagtggcag	atctctgagc	tggatgatgg	gccactacca	tctgtatatg	gctaattaaa	41700
gaccatgtgt	ggatttttta	ttcagctctt	tcgtgtcatt	cctgctatca	gcacagaacc	41760
caatctcaac	tttccagcta	tattgagcta	aacttctcac	ctcatggaat	ttgcagataa	41820
agttcaaaa	gatccttgcc	ttttcaaaat	aattttgaat	ggttgagtag	tccctctgtg	41880

28/122

ctctctcact	gacaccctct	caaggctgct	gagcacgtgc	catgctatgg	ctttctccaa	41940
catcaggaaa	tgttctccac	tcagtttcac	cttaatacaa	atgtgttctc	tcttcagaga	42000
aggcaaaaaa	attcatgacc	atctgactgg	gagaagtcac	ttctaggtaa	agtgtccatc	42060
tttttctgag	gaacacagga	ggaaaatctt	acagaaaaga	gttaacacag	caggcctaag	42120
actgcttttt	aaaataaata	aataaataaa	taaataaata	aataaataaa	taaataaata	42180
aataaatgaa	tgatagggtc	ttctgtattg	gccaggctag	tctcaaattc	ctggcttcaa	42240
gagatcctcc	caccttgggtc	tcccacagtg	ttgggattat	agacatgagc	catttgtgctt	42300
ggcccaagac	tgttattctt	aaaaagttct	ataaaaagca	tggttaatcc	ttggctggca	42360
cctgggaact	tagatttcag	aagggttccc	accatccaac	ctggaaagag	ggactcactg	42420
tgccataaatt	attgtgtggt	ttatgctgaa	ctcctgtctt	tcttcaggta	gcgtggaatg	42480
tggtatgtgc	tgggcaaaagg	gggcctgcat	gaccagcccc	caataaaaaac	cctgggtggt	42540
gggtctctag	tgagtttccc	tggtagacag	catttcacat	gcgttgtcac	agctccttcc	42600
tcggggagtt	aagcacatac	atcctgtgtg	actgcactgg	gagaggatgc	ttggaagctt	42660
tgctgtggtc	tctttgggac	ttggccccat	gcacctttcc	ctttgctgat	tgtgctttgt	42720
atcctttcac	tgtaataaat	tacagccgtg	agtaaccacc	atgctgagtc	ttccaagtga	42780
accaccagat	ctgagcatgg	tcttgggggc	ccccaacaca	gaaataaatt	ataaaagacc	42840
aaqgactggg	catggtggcc	catgccggta	atctcagcgc	tttgggaggc	cgaggcagga	42900
ggaccagtta	agcccaaaag	ttcaaagtta	cagtgacctt	tgactgcgcc	aatgcactct	42960
aaactgggag	tcttaaaatt	caatggaatg	gtagaaacat	ttttaaaaaca	ctaaatcaaa	43020
cttccaagtg	aaaacaagag	tgccgatggc	caactaaaat	gtctaggaaa	tttctgaaaa	43140
gtaaaagta	ctcagaacca	gattacctga	gcaaaccata	gccaataaca	agcttgggag	43200
taaaactgta	tgacagaagga	aatggtaaca	ggtttccagg	aacagacttg	taacagcaga	43260
taaaacagca	gaggtagaac	ctgacaaggt	gattacctgg	ggaactgcag	tctgaatgac	43320
cagactgtgt	ggacccttcc	cctcacatgg	aatacacacg	ccactcagca	gcacaccaca	43380
gctcttcaac	aatcacagga	ggcacgctac	gcctagtaag	acaggaaaaa	aggaattctc	43440
aaacttcgaa	gatgaacaca	taaagaatca	ccaagttttt	attcagtatg	atgaaacagg	43500
gacactgaat	caacagaaca	caaaccceaag	caaagataat	tactagagca	catagaagaa	43560
attattagat	attccttggga	agacctaagg	ggacattata	aagagcaagc	agttggtatg	43620
tgacgatctt	tgtgatatac	caagaaataa	aaacacagga	tgaagaccag	atagagaata	43680
atgctactat	ttgtgcaaaa	aaggagaaat	ggagaatctg	attcatatgt	gcttgtatgt	43740
gcataagaa	actttggaag	gtacataagt	aactaacaac	aatgggtacc	tacttgttaag	43800
gcgagagaa	taattgggaca	ggaatgggtg	gaacaccttt	tgtgtccgga	attggtgggt	43860
tcttgggtctg	acttggagaa	tgaagccgtg	gaccctcgcg	gtgagcgtaa	cagttcttaa	43920
aggcggtgtg	tctggagtgt	gttcccttctg	atgtttggat	gtgttcggag	tttcttccct	43980
ctgggtgggt	cgtagtctcg	ctgactcagg	agtgaagctg	cagaccttcg	cggcgagtgt	44040
tacagctctt	aaggggggcg	atctagagtt	gttcgttctc	cctgggtgag	tcgtgggtctc	44100
gctagcttca	ggagtgaagc	tgcagacctt	cgaggtgtgt	gttgacagctc	atatagacag	44160
tgacagacca	aagagtgaagc	agtaataaga	acgcattcca	aacatcaaaa	ggacaaacct	44220
tcagcagcgc	ggaatgcgac	cgcagcacgt	taccactctt	ggctcgggca	gcctgctttt	44280
attctcttat	ctggccacac	ccatatcctg	ctgattgggtc	cattttacag	agagccgact	44340
gctccatttt	acagagaacc	gattggtcca	tttttcagag	agctgattgg	tccattttga	44400
cagagtgtctg	attggtgctg	ttacaatccc	tgagctagac	acagggtgct	gactggtgta	44460
tttacaatcc	cttagctaga	cataaaggtt	ctcaagtccc	caccagactc	aggagcccag	44520
ctggcttcac	ccagtggatc	cggcatcagt	gccacagggtg	gagctgcctg	ccagtcccgc	44580
gccttcgcgc	cgcactcctc	agccctctgg	tggtcgatgg	gactgggcgc	cgtggagcag	44640
ggggtgggtgc	tgtcagggag	gctcgggccc	cacaggagcc	caggaggtgg	gggtgggtca	44700
ggcatggcg	gccgcaggtc	atgagcgctg	ccccgcaggg	aggcagctaa	ggcccagcga	44760
gaaatcgggc	acagcagctg	ctggcccagg	tgctaagccc	ctcactgcct	ggggccgttg	44820
gggcccggctg	gcggccgctg	cccagtgcgg	ggcccgcaca	gcccacgccc	accgggaact	44880
cacgctggcc	cgcaagcacc	gcgtacagcc	ccggttccc	cccgcgcctc	tccctccaca	44940
cctccctgca	aagctgaggg	agctggctcc	agccttggcc	agcccagaaa	ggggctccca	45000
cagtgcagcg	gtgggtgtaa	gggctcctca	agcgcggcca	gagtgggcac	taaggctgag	45060
gaggcaccga	gagcgagcga	ggactgccag	cacgctgtca	cctctcactt	tcatattatgc	45120
ctttttaata	cagctgggtt	ttgaacactg	attatcttac	ctattttttt	tttttttttt	45180
tgagatggag	tcgctctctg	tcgcccagac	ttgagtgacg	tggtgccatc	ctggctcact	45240
gcaagctccg	cctcccgggt	tcacaccatt	ctcctgcctc	aacctcctga	gtagctggga	45300
ctacaggcaa	tcgccaccac	gcccagctaa	ttttttatgt	tatttttttt	ttagtagaag	45360
cggagtcca	ccatgttagc	cagatgggtc	caatctcctg	acctcgtgat	ccatccgcct	45420
cggcctccca	aagtgcctggg	attacagacg	tgagccactg	cgccctgcct	atcttaccta	45480

29/122

tttcaaaagt	taaacttta	gaagtagaaa	cccgtggcca	ggcgtggtgg	ctcacgcctg	45540
taacccccagc	actttgggag	gccgaggcgg	gcggatcacg	aggtcaggag	atcgagatca	45600
tccctgggttaa	cacagtga	ccccgtcgct	actaaaaata	caaaaaatta	gccgggctg	45660
gtgggtgggca	ccggcagtc	tcgctactgg	ggaggctgag	gcaggagaat	ggcgtgaacc	45720
tgggaggcag	agcttgagc	gagccgagat	agtgccattg	ccttccagcc	tgggcgacag	45780
agcgagactc	cacctcaaaa	aaaaaaaaaa	aaaatagaga	cccggaaagt	taaaaatatg	45840
ataatcaata	tttaaaaaa	ctcaagagat	gggctaaaga	gttgacggaa	caaattctaaa	45900
tattagattg	gtgacctgca	aaaccagccc	aaggaacatc	ccagaatgca	gccccataaag	45960
ataaagagag	cattttccgct	gggcacagtg	gtatggcagg	ggaattgcct	gagtcctaaga	46020
gttgccaggtc	acattgaacc	acaccattgc	actccaggcc	tgggcaacac	agcaatactc	46080
tgtctcaaaa	aaaaaaaaaa	ttaaatttaa	aaagacagaa	tatttgagag	aaaaaaatgc	46140
ttattttcaag	aaacatgaaa	gataaatcaa	gatatttctaa	ttcccaagta	agaataattc	46200
cagaagcaga	aaatagaata	gaggcaagga	aacactcaaa	acttctccag	tgccatagaa	46260
atgtgtatta	atcttttagaa	tgaacggac	taccaaagtc	tgagcaggaa	gaacaaaaga	46320
gatccactct	taagccagtg	tggcgcccaa	gcgcagtgcc	tcagtcctgt	aatcccagca	46380
ctttgggagg	ccgaggcagg	tggatcacct	gggtcaggga	gtttgagatc	agtcaggcca	46440
acatggtgaa	accctgtctg	tactaaaaat	acaaacatta	gctgggtatg	gtggtgcaca	46500
tctgtaatcc	caactacttg	ggaggctaag	gcaggagaat	cacttgaaac	caggaggtgg	46560
aggttgtagt	gagccgagtg	catgccacac	tcccagcctg	ggtgacagag	caagattcca	46620
tctcaaaaaa	aaaatccact	cctagacaaa	taatagttaa	attttagaac	accaaggaga	46680
aagaaaaaaa	attgtaaagc	ttcagagaaa	ataaacatta	actacaaaga	aacgagagtc	46740
agacgcgtgc	acttcttcct	agataccagc	agataaagca	atatctccaa	aattcagaag	46800
gttttaacgt	agaatcctat	acccagtcac	gaatattcac	atggaaaagt	gaaataaaaa	46860
acattgttta	aacatgcaag	ggttcagaaa	gtttaccatt	cacagaatcc	ctgaaaacaa	46920
aaccaaataa	tcacttaagg	actcattaag	aaaacaaatg	aaataaaaagc	accaatgatg	46980
agtaaataat	cagaaaaatt	tacagtttac	ctaaataact	gtttatgcat	aatgtatgaa	47040
aacccaaaaa	tttaatatgg	gacagaatta	aaatcatgat	aagattcttt	tttgctttac	47100
tcattggagag	ttcacataaa	cagattatct	tttaatatgca	agagaaaaaa	atgttttagat	47160
atgtgtgaaa	aactaagggt	accaaaacag	tgcaaatcca	tttatcatca	ggaaaatcca	47220
aattaaaacc	acagtatcca	ccagaataac	taaaaggtaa	aagacagaaa	ttaccaagag	47280
ttggcaagaa	tgtggagcaa	ccacatatac	ttctggggta	aataagttgg	tgcaaccggg	47340
actgaaaact	gtttgctagt	atctactaaa	accgagcaca	tgcacagact	acaaccaagc	47400
agttccactc	actcaacaga	actcaacaga	actcactcaac	ctcactcaac	aaaagacgtg	47460
tactagagtg	ttcatgtact	tactattcat	aatagtcac	aatgcaaac	aaccaactgc	47520
caatcaaagt	caaattgtata	tctatattag	ggatatatac	aatggcatat	acacagcaat	47580
gagaatgaaa	tgaaccagct	cggcacagtg	gttcatgcct	gtaatctcag	cactttgggg	47640
gggtaaaggca	ggcagatcac	ttgaggtcag	aaatttgaga	ctagcctggc	caacacgggt	47700
aaaacctgtc	cccactaaaa	acacaaaaat	tagccgggca	tagtggttgc	aggcctgtaa	47760
ttccagctac	tcgggaggct	gggttgggag	aatcggttga	acccgaaagc	cggagggtcg	47820
agtgagcggg	gatcgtgcca	ctgcaactca	gcctggacga	tagagcaaga	ctccgtctca	47880
aaaaaggaaa	tcaaaaatat	aaaataagat	gacaggaata	atccgcaaaa	gatcagtaat	47940
caaaaataat	ataaatgggc	taaagctacc	tattaaaaga	caaagatttc	acacccataa	48000
ggatagctac	tatcaaaaaa	agagagagaa	taacagatgt	tagcaaggat	gtatggaaac	48060
tgaaattctc	acgcattgct	ggtgagaata	taaaatgggt	cagcctctgc	ggaaaacact	48120
atgctgggtc	atcaaaaaat	taaaaataga	agtactactt	gatccaacaa	ttctacttct	48180
gggtatatac	caaataaact	gaaagcaggg	tcttgaagag	atatttgtac	acccatgatc	48240
atggcagcat	tattcataat	agctatgatg	tggaaccaac	ataaatatcc	tttgataaat	48300
atatggataa	gcaaaatgtg	gtgtatacat	tcaatggaat	attaattagc	aataaaaaatg	48360
aagaaaattc	tgacacatgc	tacaacatgg	atgaaccttg	agggcattac	attaaatgaa	48420
ataagccagt	tataaaaaga	caaatactat	atgaggtact	atattagata	ctcatgcaag	48480
gtacctaata	taggcaaat	catagagaca	aaaagcagaa	tggtgggtgc	caggggctgc	48540
ggtaattggat	acagagcttc	aattttgtaa	aatgaaaaaa	ttctggagat	tggttgcata	48600
acaatgtgca	cacacttaac	actggggaac	tgtaaactta	aaagtagtaa	atggtaaaaa	48660
taaaaataat	aaataataaa	ttttatgtta	ttttaccaca	atatttatta	aaagacaaag	48720
attaactaat	taaaacaaat	ccagccataa	gctaattggt	agagtaacaa	ttaaagaaga	48780
cacagaaaat	tgaaaaatcag	tgactagaaa	aagatattcc	atataaatgc	taacaaaaag	48840
caagtacagc	aatataaaga	gaatgaacaa	aaaaaaaatt	aaataagatg	gctcgtttat	48900
tcccaaaagg	tacaattcac	caagaagata	caagaattgt	gaacctttaa	gcacataaaa	48960
cagcttcaaa	aataacaacat	ttaaagaaaa	atatatatta	aacatagaaa	tagtacaaaa	49020
accctacaa	gaatcataat	gggagtcctc	aatacaactc	tccatatcaa	caggtcaaac	49080

30/122

agagaaaaaa	aataagttaa	ggatgcagaa	aacctgaatt	accatcaata	aacttgagat	49140
taatatagaa	ctgtataccc	aatatatactaa	gagttcaggg	aacagtcgtg	actgcagtg	49200
gactgcaaat	taatctgttc	ttaatctttg	tttttctttc	agcactgtgg	cagaatagag	49260
atcctaaaaa	ccttccagct	acaaaacatc	tttttaaaaa	tataaaaaaa	tacaaaaata	49320
actctgaaat	caatagaaga	cacatggtga	aacccaaatt	ctagaatata	gggagaataa	49380
aggcattttc	agatattaca	aaaacagaaa	attgatcatt	gctgaagtaa	tttctaaaga	49440
atgtacttga	gggagaagaa	aaatgttcca	aagaaaagta	tctgtgatac	aagaaggaa	49500
ggaaagtga	gaaatggtaa	acaggtagat	aaagctaata	aatgttgacc	tagaaaaata	49560
caaaaacaat	agcaataatg	tctcggttga	aggggtgaag	taaaaatata	attaaggcca	49620
aatgtgaggt	aagtggaaatg	aaagaattag	aagtccttgc	cttggttcaca	ggactgatta	49680
aataaatgag	ccagggtttc	cattcaaaca	gttaaaaact	gaacaaaata	aactcaaatt	49740
aagtagaaag	ataaaaaaca	gaaattaatg	tcatagaaaa	ataaaaaatc	aatagaatta	49800
atcaataaat	cctgggtta	aaaagctggt	tctttgaaag	gattaataaa	ataatcatta	49860
agcaagtctg	atcaaaaaaa	aagagaaaaa	gtaccaaaaa	aagtactgta	tcagaaagag	49920
aacatcacga	tacatacaga	tatgtaagag	tctgttttct	tacaccagaa	tactatatac	49980
aacattatgc	tagcatatat	taaatttcaa	taatgttaat	gattttctag	gaaaaacaga	50040
aataattaat	ttactttgaa	gaaacagaaa	aactgagaaa	aataaatgat	catgaaaaaa	50100
atgaaaaggt	aattaaatac	tgatattaac	tgcctaatac	acaccagcag	cagcccaggc	50160
agctgcagt	caagttctgc	caaacttgag	ggaacagata	attcttctat	tccagagcat	50220
agaaaatgat	ggaaagtgtc	ccaattta	cagagaggac	agcctgatcc	ttgttatgaa	50280
cacagataaa	aatggggtaa	actatatgcc	aaactcagat	accaaacc	ttaaataagat	50340
gctagcttat	tgatgtgaac	aatccaaaag	tgcattttta	attagcccag	ggtttttagag	50400
aaagaaaatc	tagcaatgtg	accaccactt	atgttaacaa	ttttaagacg	aaaatctaca	50460
tgatcatatc	aatgcatgct	acacaaaagc	atttgggcaa	aaaacccaac	acccaccctt	50520
gactttttaa	actcttagta	attaggcata	aacagaaatg	tacttaatgt	gatagaatac	50580
actcggtaga	gatcacagagg	gaatgctccc	taaaaccaag	cccaagacaa	agattcctat	50640
ttaacctcaa	tagtcaacac	tgcagcgaga	gtaatctatg	gaagacaagg	aaaaaagtaa	50700
aaacatgaga	gacatctggt	gtttaacaga	caataagatc	acctaactgg	aagaggcaaa	50760
cgaatcaagc	gaaaaactat	taaaactgag	acaggcttta	gtatggaggc	tcagcttcag	50820
ctgtagtttg	ggctacccaa	ttcaactcgc	ttgcttgagg	agttaatcct	gcaaagctaa	50880
tttctgttga	ggtattagga	ttgacaagcc	tgtgctcctc	cctcctcccc	catcttcaac	50940
actgaaataa	cacgggtggt	ggaactggat	aacagaatct	tccaaaaaca	aaaattgtcc	51000
tgaagggtg	acttgtgccc	ttactcaaaa	aacactttat	ctgctgcctg	cagctcctac	51060
agttgctggt	ggataagcct	gccaaccagc	tcggcgtaat	tcttcctgca	gagggcaagg	51120
aagagcactt	tcacaggaaa	atttttttcc	gaactgtatg	ccgcttatta	cataaactta	51180
cgtgctggga	aatggagctc	cagcaaaaata	agatattcag	agtcaaaact	ccttaggaaa	51240
aaaaaaaaaa	aaaaagcaagc	acataaacact	aatttccttg	catgggcact	gggggaaggag	51300
gtcggtactt	ccgcagcccc	gcagggtccgc	accaccggga	aaccacggg	caccgcgcgc	51360
tgcccccg	ccttccaggt	gcactgcgcc	gcggcgcccc	agctgaccgc	ggatgcgcag	51420
ccctagccct	tccccgtca	ccccggccag	gaaggggcgg	gagcgcggcg	gacgccgagg	51480
gcgaagggtc	tctcggtcct	ctgcaccacg	cagcaccccc	aaggcacaac	agggagggtg	51540
cgggaggctc	ccgagaccca	ggagccgggg	ccgggcgtgc	ccgcgcacct	gtcccactgc	51600
ggcgagggtc	gggggtcgct	ccagggccgc	agctgtcggg	agccacctgg	ctctcagtcc	51660
cgggtccctg	cgacaacctt	cggggccgga	ggggaggagg	cggccacctg	ccgctgccac	51720
ctgcggcacc	ggtcccaccg	ctccgggccc	ggcaggacag	gccaggacgt	ccctcctggg	51780
ctggggacag	gacacgcgac	gaggggaccg	gggccccgcg	ggcgaagacg	cagcacgcct	51840
tcccagaaag	gcagtcctgt	gccccacaga	cggactgcgc	gacccccgcg	ctcgcccccc	51900
catcccttca	gaccacgcgg	ctgaggcgca	aagagccggc	cggcgggcgg	gctggcgggc	51960
cggctagtac	tcaccggccc	cgctgggtca	gcgcgcgcgc	aacccccagc	ggccacggct	52020
ccgggcgctc	actgatgctc	aggagagggg	cccgcgctcc	gccggcgctc	ccagccatcg	52080
ccgcacgggg	gccgcgcgga	gccgcgcggg	gctcgctggg	agatgtagta	ccgggacccg	52140
cgctgcgcc	gtcctccttc	agccggcgcc	cggggggccc	ctctctccca	gctctcagtg	52200
tctcatctcc	ctatctgctc	atcctctggt	cgcacataat	cgatgtttgg	gcgtcccaag	52260
ccagatgtgg	acccccattc	cgcactctac	actggagggt	ttctaagggt	ggtgccccga	52320
ccagcagctt	cagcctcatc	tgggaacttg	agaaaatgca	gattctccgt	cccacccagc	52380
ctattcgggt	tttctgcac	taaaaccatg	aaggtggggc	ccagcagtc	acattctcgc	52440
aagcccgctc	agtgatcttg	aggcgccctc	cagtttgaga	gctatgctca	cggcctcacc	52500
tccgccccgc	aaggagcccc	gtcttgccctg	tggcgctagc	cgcacacgga	cacctcatcc	52560
tgcggggccc	gccccccgcg	tgcaccctca	ccgccccacg	cctcctccgg	gatgcagcgg	52620
aggcgccctg	aagtcggcaa	ggtcaacatc	cccctcagca	tcttccttac	cctcacggct	52680

31/122

cctcctccag	gggtgcctca	tggccagggg	ttagaaagag	ccactgtgtt	tcttgacatg	52740
gaagtggcct	aagaccttaa	tgaaaactgc	aggagtggaa	tgacagaacc	tttggtcata	52800
cttgagggcg	tgaagctcaa	atgaggagga	aggaaaggat	ccagggagaa	taaccaaccc	52860
tggcaagtgt	tggcgcccag	gtagaggggc	gagcctaggc	tagcggttct	cgaccagggc	52920
cgggtgttgc	cctcctcgcc	gccccgcgta	catttgggga	gggtctggaga	catttttggt	52980
tgtcatgatg	ggggagttgc	tactgttgcc	taagtgggta	gacacgaggg	tgctcctcaa	53040
catactacct	gaaggacagg	actgccccac	aaggaagaat	gatccggccc	caaataagaa	53100
accctgggct	ggtcagcaac	aacccttttg	ttctgagaag	agaggaggaa	agaataaaaag	53160
aagtgggggt	aagtttttgg	ttggtagagg	aaacttgaag	acattttcac	tggaaaggaa	53220
gagaggaaga	ggagggagat	gtctgttaag	acgagcaaac	cgggtgacag	ctgatttcct	53280
catattgaag	taattgagtc	tagttataat	aaattcctaa	taaaaaccca	gtttatccct	53340
gcaataaaact	tgtctttttt	ttttaaatat	actgcttgat	tctgtttgct	aatattttat	53400
ttacaggcct	tgcattgata	tgcaaaaatg	agatgggcaa	taattttctt	tttgaatgtc	53460
taagtgttgt	tggtttcaga	atcaatgtta	tgctcacatc	ataaaaaatt	tggaaaccgag	53520
gcaggaggag	tgcttgaggc	cagaagttcg	agaccagctc	aggaaacaca	gtgagacccc	53580
ccatctctta	caaaaaaaaa	aaaagaaaaa	aaaatggcca	tggttgcttt	ttccttttac	53640
tctgaacaat	ttaaggagca	ttaaaattat	ctattctttg	agggttgatc	atttcccagt	53700
taaaaatgtt	cctcccagcc	tgatgctttc	tttggggagg	gtaaatcttt	taaggctaga	53760
aaagtctctt	ctgtggcaat	tttattattt	acattttaaa	aattattcta	gagttaattt	53820
tgtaaaagca	tgtattttct	aaaacaaatt	atcctttttt	tccagatgtt	caagtgtatt	53880
tgcataaagt	tgaggaaagt	agtcttttgt	gaatctttta	acttctccca	aatatcttat	53940
ttgtgtattt	tttgcttctt	tattttgtta	acttttaaaa	gtgtattttt	ttttcaaaga	54000
atcagctctt	aggtttatgt	ttttggttat	actggagctt	ttttcttctt	ctttttaaaa	54060
tattttttct	cctttatttt	ttagacgtat	tttgatctaa	cgtaatcgga	agaaggtaaa	54120
ttagaatctt	ttgttactat	tgtgttttta	tttctcttta	tttctctgaa	gtcctgcttt	54180
ataaatagta	ccatgttatt	tgtgcataaa	tattcatttg	tcttatattc	ttgggaattt	54240
tcccacttca	tcataaaaatg	accttccttg	tctcatttaa	tgtgttcaaa	ctttgccctg	54300
aatttaactt	tgtctgatat	tttaccatcc	tgtgaattt	tggttggtac	cccaaacaac	54360
ctttgctgtt	ttcgtctttt	ctgaaccttt	tattttaggt	aatcccttga	attagagcac	54420
taagltttgc	tttgtgatta	aatctgaaaa	tctttatctt	gccatagatg	agttgagccc	54480
tattcatgtg	acagctatat	tatgctgttt	catagccctt	ttggtecttt	tttactctt	54540
gcattgcata	ttttgtgttt	attgtgtttt	gtgtttcttc	tgataaattg	gaagggttgt	54600
attttttatt	agggagttgc	cttataatca	tactccgcaa	tacacatcgt	cctcagtttc	54660
ttcagactgt	ctgttaactc	cctattctga	ataaaaaatg	cattgtaatt	tccctctttt	54720
ttctttaccc	cttttcttct	cctcacctaa	tgtaaatgat	tttatccttc	tttagtattt	54780
gcttttttaa	ttaactacat	ttataaatat	ctttatcact	tgatttttaa	atcagctttg	54840
aatgagatat	ttggatttct	agatataaaa	gatgttaatt	ataccatttc	cacgttagta	54900
gggttatata	atcatacatt	ctgctgtgta	accataatcc	cacgtttgtt	ttagttccac	54960
tcctacagtt	aaaagattca	gaagtattat	taacagttat	tttgccatag	ttttttcccc	55020
aacccatttt	gtggtaagtt	atgatcctgc	tttagtttct	taagaataat	ttatagagca	55080
gagtgtgggt	gctcacgttt	gtaatcccag	cactttggga	gacaagaggt	agaaggatcg	55140
cttgaagcca	gcagttcaag	accaccttga	gcaacatagt	gagaccttgt	ctctacaaaa	55200
aatttttaaaa	tttagccaga	cgtagtggcg	tgtgcctata	gtcccagcta	ctcaggaggc	55260
tgaggcaaga	ggattgctag	agcccagaag	tttgaggctg	cagtgacctc	tgatttgtgcc	55320
actgcacccc	agtctgggca	agaaagtgag	aacctatctc	tttaaaaata	caataataac	55380
ttatgaaaaa	tatattccct	gagtttttca	tgtttaaaaa	tatttggtgc	ctttatcctg	55440
taaaagtgtt	agtataaatt	cttgggttat	actttattta	ttgaagaatg	tataagtatt	55500
gtcttctaga	attgagtgtt	gctgtaatga	aaccagaagt	cagcctgggt	tatttttctt	55560
cagaaatgag	gtaattgccc	gccggacacc	gtggctcatg	cctgtaatcc	caacactttg	55620
ggaggcccag	acaggtggat	cacgaggtca	ggagattgag	accatcctgg	ctaactgggt	55680
gaaaccccgg	ctctactaaa	agtacaaaaa	gtaggtggg	catggtgggt	gacgcctgta	55740
atcccagcta	cccgggaggc	tgaggcagga	gaatggcggt	aacctgggag	gaggagcttg	55800
cagagagctg	agatcgcgcc	actgcactcc	agcctgggcg	acagagtgag	actccgtctc	55860
aaaaaaaaca	aaaaaaaaca	aagaagtgaa	gtaattgcca	tgatgctcca	agaatttatc	55920
ctttgtctat	gaaatccaga	aatctcactg	ttatacattt	tggatttatt	attctggggc	55980
aatatttctt	gggacacaat	agattgactc	tatagcttta	attttttttt	tttttttgag	56040
acagagtctc	actgcaatct	cagcttactg	caacctctgc	ctcacggggt	caagcaattc	56100
tcctgcctca	gcctcccaag	tagctgggac	tacaggcgcg	tggcaccatg	cctggctaata	56160
ttttgtcttt	ttagtagaga	cagggtttca	ccatgttggc	caggctgggt	ttgaacgcct	56220
aacctcaagt	gatccacctg	cctcagcctc	ccaaagtgct	gggattacag	gcgtgagcca	56280

32/122

ccatgcccag	cctcaattcc	tctttctatc	tggtaatatt	tctgaagttg	aaaacatttg	56340
ttctaatac	ttatttcagt	gttcttctaa	gatgtgtaaa	gcaccctatt	cccaggctag	56400
cccccatctt	gctagtgcag	tcggctgggt	cttcacaaga	gctctgggtt	tctcctgctt	56460
aatctcaagt	acctctgtca	gcctccacct	ggtttatgat	ttggagtttt	ttgggttttg	56520
ttttttggtt	ttgacagagt	cttactctgt	cacccagggt	ggagagcagt	ggcataatct	56580
cagctcactg	caacctctgt	ctcccagggt	tgagcgattc	tcctgcctca	gcctactgag	56640
tagctgggat	tacaggcgcg	tgccaccaca	cccggctaatt	ttttgtattt	ttagttagaga	56700
tgggggtttca	ccatgttggt	cagggtgggt	ttgaactcct	gacctcaggt	aatccacctg	56760
cctcagcctc	ccaaagtgt	gagattacag	gcgtgagcca	ccgcgcctgg	catggttttg	56820
agttttaatc	tgtagtttta	ataaagatag	tgcttatggt	tgtgtttctt	atatttcttg	56880
gtactcttgg	gtaatttgta	agatcccat	atctacacaa	gaagtccatt	ttcaattctt	56940
ttcttcagac	tggttatatt	attttatatt	attttatatt	tatgtttgag	atggagtctc	57000
gctgtgtcac	ttctggaggc	tggagtgcag	tggcgcgatc	tcaggctcact	gcaacctccg	57060
tctcccgggt	tcaagcaatt	ctcctgcctc	agcctcccga	gtagctggga	ttacaggcac	57120
ctgccacttt	tttaattttt	tagagacaga	gtctcgcttt	gttgaccagg	ctggagtgcg	57180
gtgggtgcaat	catggctgac	tataacctcc	aaatcctggg	ctcaagtgat	cctcctgctg	57240
cagcctcctg	agtagctggg	actacaggca	catgccacca	tgcccagtta	attttaattt	57300
ttttgtagag	acagggtctc	catatgttgc	ccaggctggc	ctcctactcc	tggcctcaag	57360
taatcctcct	acctcagcct	cccaaattac	taggattata	agcatgagcc	accatgcccc	57420
gccttggtct	actactttta	tttcatatgt	taggtgacca	tgtaattgat	catccaaacc	57480
aggatactgt	aagaatgaaa	gaggctgaca	gtagtatgat	gctgggacta	gcatttgtga	57540
ctgagattat	ttctgggaaa	gcaggagata	cggtcacctc	acttatagtg	tgcttgtctt	57600
tggattgttg	aatattggagt	ttctatttgc	aggcttattt	caactgggca	gccttgatcc	57660
gccctgcccc	gcaatgctac	cgttctctcc	accgggtctc	tgggaccctc	tcagtcacta	57720
tacttagctc	agttccccac	cctcccactc	cctaaaagcg	taaccaggaa	tcctgcctca	57780
ggtctactgc	cgctctccgt	gggtctgttc	agttcctatt	accagagctc	aaactcccag	57840
cattccctac	ctgattccag	acttggagtc	cagagcttta	acctcttcag	gccaactccc	57900
cactttgcat	ttctgtccct	atatcttagt	ccatggagat	acatttcatt	tctttgagtc	57960
tacttacaaa	gtaaattttg	ctgtttttta	attttttttt	tgagatggag	tcttgccctg	58020
tcacccaggc	tgtggtgcaa	tgacgccatc	tcggctcact	gcaacctccg	cctcctgggt	58080
tcaagcgatt	catctgcctc	agcctcccaa	gtagctgtga	ttacagacag	gcaccaccac	58140
gccagcttaa	tttttttat	cttttagtag	agacaggggt	tcaccatggt	ggccaggctg	58200
gtcttgaaat	cctgacctcg	tgatctgccc	atctcgccct	cccaaagtgc	tgagattaca	58260
ggcgtgagcc	actgtgcccc	gccaattttg	ctttttttat	atttcattgc	tatatgttta	58320
gaggataagt	ttacagtgtc	atatgcattc	ccaaatatta	gaccaaaaaa	atctccaaaa	58380
aattagaaa	aaaatccaaa	aaatctcaaa	aaataccaaa	aagcaacaat	ctcacagacc	58440
atactcactg	acccccaaat	aaataaaatt	agaaattaac	cacaacttaa	caaaaataag	58500
tactcaagtc	agagaggaaa	gaggaaataa	acatcaaaat	tacaaagtct	aggcggtggc	58560
tcacgcctgt	aatcccagca	ctttgggagg	ccaaggcggt	cagatcacaa	ggtcaggaaat	58620
tcgagaccag	cctggccaat	atggtgaaac	cccgtttcca	ctaaaaatac	aaaaatttagc	58680
caggcatagt	gatgtgtgcc	tgtaatccag	ccacttggga	ggctgaggca	ggagaatcac	58740
tgaacccagg	gagacgaaga	ttgcagtgcg	ccaaaatcgt	gccactgcac	ttcggcctgg	58800
gtgacaaaag	gagactccat	ctcaaaaaaa	aaaaaattac	aaactcttta	gatagaaatt	58860
ttgggtgttt	tttttgagac	ggagtctcac	tctgtcgcag	aggctggagt	gcagtgggac	58920
tatgtcagct	caccgcaacc	tccatctcct	ggattcaagc	aattctcctg	tctcagcctc	58980
ccaagtagct	aggattacag	gcgccacca	ccagaccag	ctagttttta	tatttttagt	59040
agagatgggt	tttcaccatg	ttggccaggc	tggtctcaaa	ctcctgacct	caagtgtacc	59100
acctgcttca	gcctcccaaa	gtgctcagat	tacaggcggt	agccaccgca	ccccacctag	59160
atagaaattt	caacatgagg	ccgggcacaa	tggctcacgc	ctgtaatctc	agcacttcag	59220
gaggctgagg	cgtgggagga	tcaattgggc	ccaggagttc	aggaccagca	tgggtgacag	59280
agacagaccc	tgtctctatt	tatttgaaaa	aaaaaaaaaa	aaagagagag	agaaagaaat	59340
ttcaacatga	aaagtatctc	tcaaacctct	cgagatgttg	gcaaaaaagc	actcaaaagg	59400
aaatgtatta	ctgtgtgtga	atttgcttga	aaataagaaa	gaggccgggt	gtgggtggcta	59460
acacctgtaa	tcccaacact	ctgggagttc	gaatcaagtg	gatcatgagg	tcaggagatc	59520
gagaccatcc	tggttaacat	ggtgaaaccc	tgtctctact	aaaaatacaa	aaaattagct	59580
aggcgcggtg	gctcatgcct	gtaatccag	cactttggga	ggctgaggca	ggtggatcac	59640
ctgaggtcag	gggtttgaga	ccagcctggc	ctacatgggt	aaacctcgtc	tcttctacaa	59700
atacaaaaat	tagctgggag	tgggtgggtg	tgctgtgaat	cccagctact	cagaggctga	59760
ggcaggagaa	tcgcttgaac	ccgggaggcg	gaggttgctg	tgagccgaga	tcgcaccact	59820
acactccagc	ctggggcaaca	gcctgggtga	cacagtgaag	ctccatctca	aaaaatacaa	59880

33/122

aaaattagct	gggtgtggtg	gcctgcgcc	gtagtcccag	ctaccgga	ggctgaggca	59940
ggagaatgga	gtgaacctg	gaggaggagc	ttgcagttag	ccgagatccc	accactgcac	60000
tccagcctgg	gcgacagagc	aagactcttg	tctcaaaaaa	aagaaaaaaa	aaggaaaaaa	60060
gaaccctgat	aataaagaaa	ccaaatgttc	aactctcaaa	gctcggacac	tttaaagaaa	60120
taattaataa	aggcagaagt	taaagggagg	atgataaagc	aatttttttt	gttgggtttt	60180
ttgagatgga	gtcttgctct	gtcaccaggg	ctggagtgc	gtgatgcgat	cttggctcac	60240
tgcaacctct	gcctcccggg	ttcaagcaat	tctcctgcct	cagcctcctg	agtagctggt	60300
actacaggtg	cgcgccacct	ggcccagcta	atttttgtat	ttttattaga	gacgggggtt	60360
caccatattt	gttaggctgg	tctcaaactc	ctgatctcag	gtaatctgcc	cacctcggcc	60420
tctcaaagtg	ctgggattac	aggcaggcgc	caccgcgcct	ggcctaaagc	aaaatatagg	60480
ttctgtgcaa	aaggtcaata	aaaagagcaa	acgtttacaa	actggagcca	gcacccattc	60540
agctcagtg	gtctggagaa	aaaacaatct	cgcttcagaa	ttcatgatta	cgcagccctt	60600
tttgcttctt	aaaaatccta	ctatgttgct	gttgaccatt	ctctctcttt	ctctctctct	60660
tgctttctct	ccagaaaagc	tattcagaca	ttctcctctt	tcctcaaacc	ttcaaacact	60720
cctcctccat	ccttagcctc	agctgctgac	ctcacttcta	atcattgaga	aaccaggaga	60780
agcatttaag	agtgaacctc	cgctcccccg	cacgggcaaa	accaccacc	cacagaattg	60840
tgccccaat	ctgctgcctc	tcctctcacc	atggatggac	gggtccaggct	ccgagccaaa	60900
ggcaggcctc	ccctggagct	ctggatccac	cacctgcagc	ttctcaggca	gggccccagc	60960
agctcccttg	ctccctgtga	ccatcaatcc	ctccccctac	tgggtcactc	ccaacaatat	61020
atatatttag	tgatgtttct	cccatgtggt	aaaatcactt	agcctctctc	ctccccagc	61080
tactatccta	tttgtttctt	tccattctct	gcaaaacttc	tcaaagcatt	gtgtctatgt	61140
gctgactcca	tttatctctt	cccgttctct	gctgagctct	tcccacagac	tctcacccca	61200
gttactccat	gaaatgacct	ctgcactgcc	acatccaatg	gtgaatgttc	agttcttaat	61260
tttattcagt	ctttcagcag	catttgacct	ggccgatcac	tcctcttctt	taaaaatact	61320
tttctcagcc	aggcgtgatg	gctcacacct	gtaatcccaa	cactttggga	ggccaaggcg	61380
ggaggatcat	gagagcccag	gagttcaaga	tcagcctggg	caacatggca	agaccctatc	61440
tctacaaaaa	ctaaaaagta	gccagtgtga	tggcatgcac	ctgtagtccc	atctacttag	61500
gaggctgagg	cagtaggatg	acttgagcct	gggaaatcaa	ggctgcagtg	agccatgatt	61560
gcaccactgc	actccagcct	gagtgacagc	gagaccctgt	ctcaaaaaga	caaaatagga	61620
aactttttct	agcatattcc	tctgattctc	ctgctgcttc	tgtctgcaca	gattcagctc	61680
cctttgccc	ttcttctca	tcctcctgat	ctcttgacct	tgaagtggcc	cagagtacag	61740
tctttttttt	tttttttgag	acgcagctct	gtctgtcacc	caagctggag	tgcaatggcg	61800
aggtctcagc	tcagtcaacc	tctgcctcct	gggttcaagc	gatttctctg	cctcagcctc	61860
ccaagtagcc	aggactacag	gcacatgcca	ccatgcccag	caaattgttg	tatttttagt	61920
agagacaggg	ttttactata	ttggccaacgc	tggctcctcaa	ctcctgaact	cgtgaaccac	61980
cgccctcgcc	ctcccaaagt	gctgagatta	caggcatgag	ccaccacacc	cggcccagag	62040
tacagtcttt	agacggcctc	tctacctata	cttgctcccc	tcataaaactc	ctcctgcctc	62100
atggctttaa	ataccatcgg	tagactgatg	actcccatat	ttctcttttt	tttttggaga	62160
cggagtctcg	ctcagtcctc	caggtctggag	tgcagtggcg	cgatctcggc	tcactgcaag	62220
ctccacctgc	caagttcaca	ccattctcct	acctcagcct	ctccagtagc	tgggactaca	62280
ggcaccggcc	accacgcctg	gctaattttt	ttgtattttt	agtagagatg	gggtttcacc	62340
atgttagcca	ggatgggtctc	gatctcctga	cctcgatgat	cgcccatctc	ggcctcccaa	62400
agtgtctggga	ttataggtgt	gagccaccgt	gcccagccga	tgactcccat	atctctatct	62460
cttgctgtgt	gggagttctc	ctcagaactc	catactcata	aatccaactc	tcataaatag	62520
tatctcaaat	gggcaatatg	ctcaaaaagtc	aattcctact	tttctcccta	aacttgcttt	62580
cctgcagtct	ccaccatctt	aatgtccaat	ctaaccattag	gaggcaaaaa	ctttgaagtc	62640
attcttgact	cttctctatt	acacacccta	tccaatcttt	ctgcagatcc	agtcgacccc	62700
caaatccagt	tagctctcat	catctccctt	gttaccctct	gggtccaggcc	atcttctctt	62760
ctcacctgaa	tcactgcagc	attctcctca	ctgggtctct	tgggtctgtt	ttcactccac	62820
cttagcatag	tctccacaga	gcagtgcagag	ggatcctttt	aaagtgtaat	tcccatcctg	62880
tccttgctct	gctcaaaacc	ctgtcgtgat	tcccgtttta	atctgtcaga	ttaaaagcca	62940
gagtctttcc	agtgccttac	atgatctgcc	tattatcacc	tcccacttct	ttcccttgcc	63000
tcactccact	ccagctctgc	agctgtcctt	tctgtttcct	gaacagccca	gattttgctt	63060
ctttagaacc	tttgattttg	ctgtcccttc	tgtctggaat	gtttttccag	gaagtcacct	63120
ggctctctcc	tgcacttcct	tcctgaccac	cactgtttaa	aatcactcaa	acacacttca	63180
ggccggacat	ggtggctcac	gcctgtaatc	ccagcacttt	gggaggccaa	ggtgggtgga	63240
tcacctgagg	tcaggagttc	gagaccagcc	tggccaacat	ggtgaaactt	cgtctctact	63300
acaaatacaa	atagtagcca	ggtgtagtgg	cacacacctg	taatctcagc	tactcaggag	63360
gctgaggcag	gagaatcgct	tgaaccacga	aggcagagga	ggtgcagtga	gccaagatca	63420
cgccacaaca	ccccagcctg	ggtgacagag	caagacccca	tctcaaaaaa	aaaaaaagaa	63480

34/122

aaaaaaatca	cacaaacaca	cttctcttca	tattcctttt	ccaagtttta	tttttctcca	63540
gaatacttta	cattgtttta	atggaagttc	tccgtttccc	cccaactaga	atggatactt	63600
cctgcaggta	ggcactctag	tcctcccatc	caagtactaa	ccaggctcaa	ccctgcttag	63660
cttctgagag	caggggagat	caggcctggt	caggggtggt	tggcccagga	attttgattc	63720
tgttttattc	attgctgttc	tggtgattct	cttttgttcc	tcctcctagt	gctgagaaca	63780
ctacttgtag	ataataagca	ttcaataaat	atttgttgaa	tgaatgactt	gttgaatgaa	63840
ttaatctcag	aaatgcagga	ctggttctac	attagaaaa	ttttcaaggt	cattctctgt	63900
tgctcgtaaca	cattaagaga	ggaaaaat	gtactctaaa	tcatttgata	aaatacatat	63960
tgatttctgt	tttcaaaaac	tcttagtggt	tgggcgaggt	ggctcacatc	tataatccca	64020
gcattttggg	aggacgaggt	gggcggatca	cttgagggtca	ggagtttgag	accagcctgg	64080
ccatcatggt	gaaaccctat	ctctactgaa	ttaaactttaa	ttagccgggt	gtggtggcgc	64140
atgcctgtag	tcccagctac	ctgggagggt	gaggcaggag	aatggcctga	accggggagg	64200
cggagggtgc	agtgaagcaa	gatcatgcca	ttgcactcca	gcctgggtaa	cagagtgaga	64260
ctccatctca	aaagaaaact	cttagtgagt	ttaggaatcc	aaggaagacc	ctcaaaactaa	64320
atagatatac	tagctaccag	aagccttcag	taaaccttcaa	cactccatgg	tgaacatta	64380
gaaacattcc	tactaaaaga	caggctaaga	atgcctgcaa	tcttcacggc	tagtccaaga	64440
agtcaaaaag	aagaaatgag	cgctgattta	aaaaaataaa	caaacaaaaa	actaccgatg	64500
cagaggtctg	cagcaaggac	tgaaggactg	tacagtactt	gcctggagca	ggcggatggc	64560
cacaccctcg	cgaagcctgc	tcagctgggt	gggggacgct	ccagtgtgtg	agtggcagga	64620
tgaggggtac	ttcctctgct	agggagttgc	actggggaga	tcctccccc	ctcacacttt	64680
ggcagctggg	gctttggaat	gtgacttagc	ttctgtcaaa	gggtcaatcc	accctttgat	64740
atatgatgca	aaggcgaaca	tatgatgcaa	aggtgagaga	acagcccaaa	ttaggacttt	64800
taccacagct	gtggaggtgg	acagcgacag	tgggtggccc	tggccagact	tttcatgctc	64860
aaaggtgggt	gttgttcttc	ctacttcttg	tcctccagag	gcttcccttg	cctgtgtgct	64920
gaacctgctt	cttttaattt	tttttaactt	ttttaaat	ttaattgttt	taattaaaac	64980
aaattttgaa	aactgtctga	acctgctttt	gaacctgct	atgatttgaa	tgtttgtccc	65040
ctgccaaact	gattttgaaa	cttaatctcc	aaagtggcaa	tattgagatg	gggctttaag	65100
cagtgaactg	atcatgagag	ctctgacctc	atgagtggat	taatggatta	atgagtgtgc	65160
atgggagtg	catcagtggt	tttataagag	gaagaattaa	gacctgagct	agcatggctg	65220
ccccttcacc	atgtgatata	ttacactgcc	taggggctct	gcagagagtc	cccaccaaca	65280
agaaggctct	caccagatata	agctcctcaa	ccttgtactt	ctcagcctct	gtaactgtaa	65340
gaaataaaat	ccttttcttt	atgaattacc	cagtttcaga	tattctgtta	taaacaatag	65400
aaaacgaact	aaggcaaaact	ctcatgattc	tactgccatg	ccattccaat	aaactccctt	65460
tatgcttaag	agagccagag	ttggccaggc	gtggtgactc	acgcctgtaa	ttccagcact	65520
ttgggaggcc	gaggcagggt	gatcacaagg	tcaggagatc	gagaccatcc	tggctaacac	65580
ggtgaaaccc	cgctctctact	aaaaatacaa	aaaaattagc	tgggcgtggt	agtgggtgcc	65640
tgtagtccca	gctactcggg	aggctgaagc	aggaggagaa	tggcgtggac	ccaggaggcg	65700
gagcttgtag	tgagtcgaga	tcgtgccact	gcactccagc	ctgggtgaca	gaatgagact	65760
ccgtctcaaa	aaaaaagaga	gccagagttt	atctctgttg	cttgcaacca	agaaatctgg	65820
ctggtgcaat	gaagtttcca	taaataatag	caattttaaag	actctttcca	agccaggcaa	65880
tgcttagcct	tggtgagtcc	ttgtggtaat	acattcattc	attcatttgt	tcaaccaact	65940
gtgctccaga	gactaagaat	acaaaaatgg	gggcccgggtg	tgggtggctca	cacctataat	66000
cctagcactt	tgggaggccg	aggcaggtag	atcacctgag	gtcaggagtt	cgagaccaac	66060
ctggccaaaa	tggtgaaacc	cctactctac	taaaaataca	aaaaattagc	tgggggtggt	66120
ggcggacacc	tgtaatccca	gctactcgtg	agactgaggt	aggagaatca	cttgaacccg	66180
ggaggcagag	gttgcaagtga	gccgagatcg	caccactgca	ctccagcctg	ggcaacaaga	66240
gcgaaactcc	acctcgaaaa	aaaaaaaaaa	aaaaaaagag	ggccgggggt	gggcgcagtg	66300
gctcacgcct	gtaatccag	cactctggga	ggccaaggca	ggagaattac	gaggtcagca	66360
gactcgagacc	agcctgacca	acatggtgaa	accccatctc	tactaaaaat	acaaaaatta	66420
tccgggcgtg	gtggcgacac	cctctagtcc	cagctcactg	ggaggctgag	gcaggagaat	66480
cgcttgaaac	cgggaggcag	agggtgcagt	gagccgaaat	catgccactg	cactccagcc	66540
tgggtgacag	agtgaagact	cgctctcaaaa	aaaaaataaa	aaaaaataaa	gaattcaaaa	66600
attgtagagt	tatagtgtgc	ttctagttaa	gttgagagga	catctgtcct	tcaaggaagg	66660
ctagaatcta	taccctgagt	ccttactgaa	atcaatccag	cagtcaaaac	atgggaccaa	66720
cgatcacagc	agtaagatag	gaagagcacc	tttgtacatt	tagctcatgt	tgagataagc	66780
cactgacaga	gctgaaggaa	gctcacagtt	ctgggttcca	tcctttggca	tttaaaaaga	66840
aaagtgttaa	gaaaattcgg	ttggtcacgg	tggctcacgc	ctgtaatccc	aacactttga	66900
gaggccaagg	caggcagatc	acgaggtcag	gagttcgaaa	ccagcctggc	caacatgggtg	66960
aaaccccgct	tctactaaaa	acagaaaaat	tagccgggca	tgggtggcgca	tgcttataat	67020
cccagctact	caggaggctg	aggcaggaga	attgcttgaa	cccgggagggg	ggagggtgca	67080

35/122

gcgagtgaga	gcaggccact	gcactccagc	ctgggagaca	gagcaagact	ctgtctcaaa	67140
aaaaaaaaag	aaaaaaaaagaa	agaaaggaaa	aaaagaaaaga	aaaaaaaaaga	aaaaaagaaaa	67200
ttcaggccag	gccaggcctg	gtggctcaca	cctgtaatcc	caacactttg	ggaggctgaa	67260
gcgagacggt	gccttagccc	aggagtttga	gaccagcctg	agcaacatag	cgagaccctg	67320
tctctataaa	aaaaaat ttt	tttttgcca	gacgcagtg	ctcacgcctg	taatcccagc	67380
actttgggag	gccgaggcag	gtggatcacg	aggtcaggag	atggagacca	tcctgggctaa	67440
cacggtgaaa	ccccatctct	actaaaaaat	acaaaaaatt	aaccgggctg	ggtggcgggc	67500
gcctgtagtc	ccagctactc	gggaggctga	ggcaggagaa	tggcgtgaac	ccgggaggcg	67560
gagcttgtag	tgagcccgaga	ttgcgcact	gcactccaga	ctgggagaga	gtgagactcc	67620
gtctcaaaaa	aaaaaaaaaa	aaaaaaaaaat	taattgtcag	gtgtgtctgg	atgcagctgt	67680
agtcttagct	actcggggagg	ctgagggtaa	aagatcgctt	gagcccaggga	gttcaaggct	67740
gcagtaatag	tgcctctcac	tctaccctgg	gtgacaatga	gaccctctct	caaaaagaaa	67800
gaaaaaagg	aaagaagaaa	agaaagaaa	aaagagaaga	aaggaaaggaa	gaaagaaaga	67860
aaaagaaaag	gaaggaaagga	agaagaaaaa	aaaagaaaaga	aagaaaagag	agagaagttc	67920
aaagacaaaa	gggtcaggat	cccaaaatag	ttttatgctt	ttatttattt	atttacttat	67980
ttatttttga	gacagtagtg	ctctgtcgcc	caggctggag	tgcagtgatg	cgatttgcgc	68040
tcactgcagc	ctccaaactg	ggctcagggt	gccctccac	ctcagcctcc	cgagtagctg	68100
ggaccacagg	cgcgtgccac	catgccagc	taatttttta	attctttgta	gagatgaggt	68160
ctctatatgc	tgccccaggct	ggctctcgagc	tcctgggctt	aagccatcca	cccgcctggg	68220
cctcccaag	tgtgggatt	acagaagtga	gccaccgcgc	ctaategggt	ggtttgtttg	68280
tttattgacg	gggtctcgct	gctgcccagg	ctggagtggc	agtggctgtt	cacaggtgca	68340
gtcctggagc	attgcatcag	ctcttgggct	ctagcgatcc	tccagagtag	ctgcagctgg	68400
gattccaggc	gcgccaccgc	gcggggctca	gaatgggttt	ttatattgag	ggttatgctg	68460
ccacctagag	gatatatgta	gtaccgaact	gtgtgcgcag	ggaggctgag	gttgagtgca	68520
gccaagatga	tgccagggca	ctccagcggt	gggtgacagag	caagatttca	tctcaaaaaa	68580
aaaaaaaaaa	aaaaaaaaaa	aagaattgaa	agtaaggctc	tgaagagata	tttgtgcctg	68640
tatggtcata	gcagtatata	ctttgaccca	ctagctaaaa	cacaaaagca	acatgtgtct	68700
gtcagcaggt	gaacggataa	acaaaatgtg	gtatatatgt	acaattgaat	attattcagc	68760
ctttaaaaag	gaataaaaagg	ctggatgcgg	gggctcacgc	ctgtaatcct	aacactttgg	68820
gagactgagg	tgggtggatc	acccgaggtt	aggagtgtga	gaacagcctg	gccaacatgg	68880
tgaaacttca	tctctactaa	aaatactaaa	attagccggg	catgggtggca	cttgtctgta	68940
atccaagcta	ctggggaggc	taaggcagga	gaattgcttg	aactcaggag	ccggagggtg	69000
cagttagcta	actgggcacc	actgcactcc	agcctgggca	acagagtggg	actccatctc	69060
aaaacaaaca	aacaaaaaat	tattattttcc	aaagaaacaa	gaccctgggt	ccattttcca	69120
gcccacacct	gatgttgact	cacaacacac	agcctgggtt	gctatgagcc	tgcttcattt	69180
aattgtcacc	ttaacttcac	atcacctcca	agtcctggaa	taactctttg	ctgacctttg	69240
tgtgctgagc	catctccatg	tcgctcaacg	tgcagtcctt	ctcactgcac	tgagtcaata	69300
gccagacgtg	ctctgactgc	agggtcatcc	ttgtggctt	aggctgactc	gggcatagca	69360
gggtgctctg	agacctcacc	gcatataggc	tttgcccca	ataaactcta	tataatattc	69420
atattatgtg	gtctgggtgt	gtgtagcttt	gcactgtctt	ctcgtgacag	tgccctcaac	69480
ctctttccca	ggatttctct	ctctacctcc	tcaagtccca	ctgctctgca	aagacaaaaa	69540
gctgcagagt	cccagctccc	tcctttacac	cccacgacgc	agcctcctct	ctcagaaccc	69600
tttaaacaga	gtcttttact	gcagatccca	agaacagcca	caccctctct	tcccaccacc	69660
tccagacaca	cccaggtaat	tatagcacc	agggttaacta	tgtagatgga	gtccctggaa	69720
catgtggata	gtgccccctg	ggagtatgca	aaagcaacat	tgctggcacc	tgagagagac	69780
agggtgacat	ccaggaatca	gagcatgggc	ctctgggagg	tagggatgtg	gccaggcagg	69840
ctgccaaaaa	ttggtagagc	aaggccacag	gatctttctg	accttccttc	caaacagagg	69900
ctcctgtact	ggtgatccct	gtgttgattg	accactccct	tcctgggggt	cgtggtctct	69960
gtcccagttg	cccggacttc	tgtgagtgtc	ctactgaggt	ccttttcatg	agaagcatgc	70020
tgtccttcca	cctgctggga	gcaagagtga	caacttcaat	actataatag	cagtggcata	70080
cagagaagaa	gaaagatgaa	gtggcaagaa	aaacaggctt	ccaagcagga	gtttttctat	70140
aaaaacaaaa	acgtttacaa	gcaaaat ttt	tataaagggc	tagatagtaa	atattttagg	70200
ctttgagagc	cacatagact	tgtttgacag	gactcaatgt	cgctattgta	gtttgaaagc	70260
agccatcagg	gttatgtaaa	tgagttagtc	tgattttgtt	tcagcaaaat	tttattttacc	70320
aaaacagaca	atgagtgggc	tggatttggc	ccatgatctt	tagtttgcca	actcctgctt	70380
tgggctcacc	cagatctgat	tttgaattct	ggctctgcta	ctggtagctt	gcaggagctt	70440
ggaaggctct	ctgagcctgt	ttcctcatct	gtaaaattaa	agcaataatt	tctaaccact	70500
aagagtgtta	cctcacgcct	gtaatcccag	cactttggag	gctgaggcag	gcggatcacc	70560
tgaggctcaga	agt tcaagac	cagcgtggcc	aacgtggcaa	aaccctgtct	ctactaaaaa	70620
atacaaaaaag	tagccgggca	tgggtggcgc	catctgtaat	cccagctact	tgggaggctg	70680

36/122

aggcagggat	actgctagaa	cctgggaggt	ggagcgtgca	gtgagtggag	atcacacctc	70740
cacactccag	cctggccgac	agagcggagac	tccatctcaa	aaaaaaaaaa	aaaaagagt	70800
ttagaagggt	ttgagataat	gaataaaaaga	tgccttggtg	ataactaagta	ttcaacaact	70860
gatagctgca	ttggtctaata	tataacaggt	tagaagcgat	tgagtcaaca	aatgctggat	70920
ttgtcagggg	ggacttccta	tcaggaggta	gatcttgggc	tgagtcctga	agcaaagata	70980
ggcattggat	agaggagttg	agagaacacc	ctaggactgt	tattattatt	attcgacacg	71040
gagtcctctg	ctctgtcacc	caggctggag	tgcagtggcg	cgatctcggc	tactgcaaac	71100
ctctgcctcc	cagggttcaag	cgattctcct	gcctcctaag	tagctgagac	tacagggtgtg	71160
tgccaccaca	cccggctaata	ttttatattt	ttagtagaga	cagagtttca	ccatgttggc	71220
catgctgggt	tcgaactcct	gacttcaggt	gatccaccgc	cctcagcctc	ccaaagtgtc	71280
ggaataacag	atgtgagcca	ccgcacccag	cccagaacca	tttttcaatc	cttggctctg	71340
ccttttatta	gctgcaagat	ctcaggcaat	ttatttaacc	tctccaaaga	ctcattttct	71400
cattcacaaa	atgaggcaaa	taataatatc	tactatccca	ggttgtcatg	agaattaaat	71460
gcaacatgac	atttaatgaa	atgagaagtc	ccttggaact	taactggcta	aagtatgtgc	71520
tcgacaagga	tatcatttta	ggtaggatact	tagcatctca	gaactgatgc	tcacaatgga	71580
atatcattga	aacgcattaa	aattcatttt	aaatgattgt	aggtagtgag	gcaattgaaa	71640
gaagaagaca	agaggactga	ttataatgct	tcaggctcac	tagtctcctt	ttaggaggga	71700
aaaacaattt	caagttaaat	tttaggctct	agatttttct	ccctgctgct	cattagaatc	71760
accacagatt	atgaaatcag	agcctctctg	agctgtgtt	tttcatctcc	agaatgagag	71820
ctggttggtg	gattaaagttt	ttgaaaaagt	acatctgaac	ggtagatcgaa	aatgatagtg	71880
atattattgc	agtgatgggt	attattgttg	ttattattat	actgaaagag	gcttcagttt	71940
tctgatccat	aaagtgaggg	aattgcatga	gaccattgct	aagattcctt	ctagctctgt	72000
ttttttgttt	ttgttttttt	gacagagtct	ctgtcgcccc	ggctggagtg	caatggcatg	72060
atcttgggtc	actgcaacct	ccgcctcccc	ggttcaaatg	atcctcctgt	ctcagcctcc	72120
gaagtagctg	ggactacagg	cacacaccac	catgcccagc	taacttttat	atttttaata	72180
gaggtggggg	ttcaccatata	tggtcagggt	gggtctcaaac	tcctgacctc	aggtgatcca	72240
cccgcctcgg	cctcccaaca	tgctgggatt	acaggcatga	gccactgtgc	ccaacccctt	72300
ctagctttct	tgatcactga	ttctagggtt	ctctgctgaa	atatatttga	gacatcctgg	72360
ataaaagatc	atgcaagagc	tcccaatatg	gtattaataa	ttgattctgg	aggcttagct	72420
actcctgatg	gattagacat	gactcaactg	cctctcttat	gtgtacaaca	caacaacaca	72480
accaagaaa	gttattcttg	cattccattt	attcagttta	tttacagccc	ttacttccag	72540
cagcacgtta	actgatggc	caggggccgg	tgcagtggct	caagtctgta	atcccaggac	72600
tttggggagg	caagggtggc	ggatcacaa	gtcaggagtt	tgagaatctg	gcaattcttc	72660
agacttagaa	gcaaccagct	cgataacaca	gtcttggtg	ggctctccct	ctgtccctcc	72720
ctcgcttccc	tcatttctca	tccttgcccc	tgagactgtg	caccttcaca	tagccctgcc	72780
atgagacctt	catctcaggc	tttgctttct	ggggtaactg	aggctaataa	ctgagtggcc	72840
ctaaaagagg	attgggattt	ggaagttaga	ttattcacca	gagaacagac	tttgctgatg	72900
atcaggccca	gggtgttaatt	gttgaaaaaa	agagaggatg	catagtctta	tctcatctcc	72960
tagtcaaagt	caacaccatg	ataaataaga	gtcaaatcct	gagatgtgaa	ttggggacat	73020
ttgagtgggt	aaccttgaga	agcttgcacc	ttcagacccc	tcaatacccc	tgctccccag	73080
agaaggcttg	acattgacct	cagcacaggc	aggagccctg	caagatgcca	tttgtcctac	73140
taaagatgga	cccctccact	ctgtttctag	gtaaataaacc	aaagtcaagt	ctccacacag	73200
cctgagcaag	aaagtccagag	cctgctacag	gagaaaatac	cacactggcc	aaaggattca	73260
ctagccctgg	ccactgtgtg	tgggaggaac	cagggaatca	tgtgtgggag	tcaatgttga	73320
agctgttgga	ctgggggtgg	gggtggaatat	aagcctggcc	ctggggaggt	tttcccgttt	73380
gagggccttt	accacacaact	caagatccag	tgctatagca	ggagatccca	gagctagtcc	73440
taacagatgg	tcaggattga	acttggccta	gagtaaaatg	aggaggatag	tgccagaact	73500
ttctcaacat	actattgagg	aagaggtcag	aaggcttaag	gaggtagtgt	aactggaaag	73560
gggtcctgat	ccagacccca	ggagagggtt	cttggacctt	gcataagaaa	gagttcgaga	73620
cgagtccacc	cagttaaagt	aaagcaattt	tattaaagaa	gaaacagaaa	aatggctact	73680
ccatagagca	gcgacatggg	ctgcttaact	gagtgttctt	atgattattt	cttgattcta	73740
tgctaaacaa	aggggtgatt	atttgtgagg	tttccaggaa	aggggagagg	atttcccaga	73800
actgatggat	ccccccactt	ttagaccata	tagagtaact	tcctgacgtt	gccatggcgt	73860
ttgtaaaact	tcatggccct	ggaggggaatg	tcttttagca	tgtaaatgta	ttataatgtg	73920
tataatgagc	agtgaggacg	gccagaggctc	gctttcatca	ccatcttggt	tttgggtggg	73980
tttggccggc	ttctttatca	catcctggtt	tatgagcagg	gtctttatga	cctataactt	74040
ctcctgccga	cctcctatct	cctcctgtga	ctaagaatgc	agcctagcag	gtctcagcct	74100
cattttacca	tggagtgcgt	ctgattccaa	tgctctgac	agcaggaatg	ttggaattga	74160
attactatgc	aagacctgag	aagccattgg	aggacacagc	cttcattagg	acactggcat	74220
ctgtgacagg	ctgggtgggt	gtaattgtct	gttggccagt	gtggactgtg	ggagatgcta	74280

37/122

ctactgtaag	atatgacaag	gtttctcttc	aaacaggctg	atccgcttct	tattctctaa	74340
ttccaagtac	cacccccgc	ctttctcttc	ctttctcttc	tttctgattt	tactacatgc	74400
ccaggcatgc	tacggcccca	gctcacattc	ctttctctat	ttaaaaatgg	actggggctg	74460
ggcgcggtgg	ctcatgcctg	taatcccagc	actttgggag	gccgaggcgg	gcggatcatg	74520
aggtcaggag	atcgagacca	tcctggctaa	cacggtgaaa	ccccgtctct	actaaaaatg	74580
caaaaacatt	agccaggcgt	ggttgcagggt	gcctgcagtc	ccagcggtct	aggaggctga	74640
ggcaggagaa	tggcgtgaac	ctgggagggtg	gaggttgcaa	tgagccgaga	ttgtgccact	74700
gcactccagc	ctgggtgaca	gagcgagact	ccgtctcaaa	aaaaaaaaaa	aaaaaaaaaa	74760
tagctgggca	tgggtggcgcg	tgcctgtaat	accagctact	ctggaggctg	aggcaagaga	74820
atcgcttgaa	cccagtaggc	ggaagttaga	gtgagccgag	atcttgacac	tgcaactccag	74880
cctggtgaca	gagtgagact	ctgtctcaaa	aaaaaaaaaa	agaaaaaaag	agacagaaag	74940
aaagagcaca	gacagagtca	cagggtatttg	cagtaggaag	ctgtcagggtt	agagtgcacg	75000
gaaatagaaa	gtatatttta	cacttacagc	acatcttctg	ttgatttagcc	acatttaaaa	75060
tactgaatgc	caacgtgtgg	ctatttagta	ttcactaaaa	tcttgagacag	tgcaagtcta	75120
aagaatcctt	gaatcgctcg	gcatgggtggc	tcacgccttt	aatcccagca	ctttgggagg	75180
ccaaggtgga	aggatcactt	aagggtcagga	gttcgagacc	agcctggcca	acatgggtgaa	75240
acctcgtctc	tactaataat	acaaaaaaaa	ttagccgggc	atgggtggtgc	atgcctgtaa	75300
tcccaggtag	ttgggaggct	gaggcaggag	aatagcttga	atccaggagg	cgctgcagtg	75360
agccgagatc	atgccatgcc	actactgcac	tcacgcctgg	gcaacagagt	gagactgtct	75420
caaaaaaa	aaaaaaattg	ttgggcgtgg	tgggtcacgc	ctgtaatccc	agcaactttg	75480
gaggctgagg	ggggtggatc	acctgggttc	tggagttcga	gaccagcctg	gccaacatgg	75540
tgaaacccca	tctctactaa	aaatacaaaa	attagctggg	cgtgggtggtg	ggcacctgaa	75600
atctcagcta	ctcaggaggc	tgaggcagga	gaatttcttg	aacccaggag	gcagagggtg	75660
cagtgcagcca	agatcgcgcc	tctgcactcc	atcctgggtg	gcagagcaag	actatgtctc	75720
aaaaaaaaaa	aaaaaaatac	ttgattgtct	ggacattctg	cagaacatca	tatggagaca	75780
ctatgttgac	gacatcatgc	tgattgtgaa	caagaaatgg	caagtgttcc	agaaacacag	75840
tcaagacaca	tacatgccag	aagggtgagat	ataaactcta	ctaagattca	gtggcctgcc	75900
acactgggtga	catttttaaa	cctgctagat	gtttgtgtag	aaaaggattt	aaccttgccc	75960
aaagaggggt	ctggcctttg	tccccageta	ctggacataa	tctctttaaa	ctcttgaaat	76020
atcatctctg	atagaagtat	ttttgttttg	actaggggcc	ttgggcccagc	cagatagcaa	76080
caatgtgatc	tgggttgggg	gctttggatc	aggtggcatc	agtgtgacct	cctgagtggc	76140
tagagactag	atcaaacac	atgggcagac	aacccagctt	acatgatgga	attccaataa	76200
agactttgga	cacaaggcct	tgggtaagct	ttcctgggtg	gcaatgctct	atactgggaa	76260
acccattctg	actccatagg	gagaggacaa	ctggatattc	tcatttggtg	cctccctggg	76320
ctttgcccct	tgcatttttc	ccttgtctga	ttattattat	tattatgaga	tggaatctcg	76380
ctctgtcacc	caggctggag	tgcatgtgaa	tgatctcaac	tcactgcaac	ctctgcctcc	76440
ccgggtcaag	cgattttcct	gtctcgccct	cccgatagc	tgggactaca	gatgcatacc	76500
accacacccg	gctaattttt	ttgtattttt	agtagagacg	gggtttcacg	ttagccagga	76560
tgggtctcgat	ctcctgacct	catgttccgc	ctgcctcggc	ctctcaaagt	gctaggaata	76620
catgtgtgag	ccaccgcgcc	cagccccctt	ggctgattat	taaagtgtat	ccttgagctg	76680
tagtaaatat	tataacagct	tataacagct	tttagtgagt	ttgttgagca	cttctagcaa	76740
attatcaaac	ctaaggatag	ccttggggac	ccctgaactt	gcagttgggtg	tcagaaataa	76800
gggtgctcat	gtgtgtacca	tgccctctaa	ttttgtagtt	aattaaacttt	cacaacttta	76860
ttattaccgc	ttacactcaa	tggtttattca	catttatcca	cataccactt	attctagtgc	76920
cttgcatcaa	agactttcta	tctcatgtac	tttattctgc	ttgaagtaaa	tccttttagga	76980
tattcttttt	tttttttaaa	ctttgcacat	acatactttt	attttttatt	tatttttaat	77040
tttgttattt	ttgtgggtac	gtagtagata	tatgtattta	tggagtacat	gagatgtttt	77100
gatacaggca	tgcaatgtga	aataagcaca	tcatggagaa	tggggtatcc	atcctctcaa	77160
gcaattttat	cttcaagtta	caaacaatcc	aattacactc	tttaagttat	tttaaaatgt	77220
acattttaatt	ttgtattgac	tagagtcact	ctgttgtgct	atcaaataa	attttttttt	77280
tttttgagac	agagtctcac	tcagtggccc	agactgaaag	tgcaagtggc	caagctcggc	77340
tcacttcaat	ctctgcctcc	ctgggtcaag	cgaatctcct	gcctcagcct	cccacatagc	77400
tgggattaca	ggcacacacc	accatgccc	gctaattttt	atattttttt	agtagagacg	77460
ggttttcgcc	atgttgggca	ggctgggtctt	gaactcctgg	cctcaaataa	tctgaccacc	77520
tcagcctccc	aaagtgcctag	gattacaggg	atgagccacc	acacctggcc	aaaatagaat	77580
attcttttagt	gaggtctgct	gggtgacaatt	ttttctttt	ttttgagact	gagtcctgct	77640
gttgctcagct	tgggctggag	tgcaatagca	cgatctcagc	tcactgcaac	ctccacctcc	77700
cggattccag	caattctcct	gcctcagcct	cccaagtagc	tgagagatta	caggcaccca	77760
ccaccacacg	cggctaattt	ttgtattttt	agtagaaatg	ggggttcacc	gtgttgggca	77820
ggctgggtctc	gaactcctga	cctcagggtga	tccaccacc	ttggcctccc	aaagtgcctg	77880

38/122

gattacaagc	atgagccacc	acgcacagcc	aatttttttc	gtttttgtct	gaaatcttat	77940
tttgtgtcat	ctttgaaata	tattttttgat	ggatataaaa	ttgttggttg	atagttatta	78000
tcaitattat	tattattttg	agacaggggtc	tcactctgtt	gcctatgctg	gggtgtagta	78060
atgtgatctc	ggttcactgc	agacttgacc	tcctagggct	caggtgatct	tcccacctca	78120
gcctccctag	tagctgggac	tacagatgca	tgccaccata	cccaactaat	ttttctatct	78180
tttgtagaga	tgaggctttg	ccacatttcc	caggctggtc	tctaactcct	gagctctagc	78240
aatccacca	ccttggcctt	acaaagtgtc	gggccatgac	tagccagcag	ttacttttta	78300
tagcatattg	aatattttaat	atgaatcttc	tggcatccac	tgtaaactgtt	taaaaaatca	78360
gctgttact	tggcactctt	tttttttttt	ttttttttga	gacagagtct	tgccctgtcg	78420
cccaggctgg	agtgcagtgg	cgtgatcttg	gctcactgca	agctctgcct	cccgggttca	78480
cggcattctc	ctgcctcagc	ctccggagta	gctggggact	aaggcgcccg	ccaccacgcc	78540
cggctgattt	ttttgtattt	ttcgtagagt	tgggggttca	ccgtgttagc	caggatgggtc	78600
tcatctctct	gacctcgtga	tctgtccgcc	tccggcctccc	aaagtgtctg	gatttataggc	78660
gtgagggacc	gcgcccagcc	tctttttttt	tttttttttag	acggagtctt	actctgtcat	78720
ctaggcttgt	gtacagtggc	gtgatctcag	ctcagtgtca	cctccacctc	ctgcctcagc	78780
ctgcaataa	gctgggatta	caggtgcgta	ccatcacgcc	cggctaattt	ttgtattttc	78840
agtagatag	gggttttcacc	atgttagaca	ggctgggtctc	gaactcctgg	cctcaagtga	78900
tcttccctgc	ccagcctccc	aaagattaca	ggcctgagcc	accgcacccg	gccaaagtgc	78960
actcctttga	aggtaacttg	cttcccctac	ccctagcaat	ttttaacaat	ttttcttcat	79020
ttttatttcc	tgaagttttg	ttattaataa	tctgtgtgca	gatttctttg	tatttctttt	79080
gtttgttgtt	catagtgtat	cttgaattag	tgtgttggtt	tctgttatca	ccacaggaaa	79140
attgtcagcc	gttagctttt	caaatatttc	cttgctaaat	tctctcttct	cccctttcgg	79200
tacuatat	ttgattaaaa	ctaaaaccag	ggccgggtgc	agtgaactcat	gcctgtaatc	79260
ccaacacttt	gagaggctga	ggcaggtgga	tcacctaaagc	tcaggagtgc	aagaccagcc	79320
tggccaatat	ggtgaaaacc	cgtctctact	aaaaatacaa	aaattaccag	gcatgggtggc	79380
acacatttgt	agtcaggagg	ctgaggcagg	agaattgctt	gaatccagga	ggtggagggtt	79440
gcagtqagct	gagatcccac	cactgcagtc	tggcctgggc	gacagagtga	gatgagaatc	79500
tgtctcgaaa	aaaaaagtta	tgaatgtttg	ataaactata	tttgttagaa	tgtttgttgt	79560
agaatactat	tcattgtattt	ttaaacaatg	ttagattaaa	ccatttactg	gattttgtgat	79620
aattaaacta	ctgattttac	ctcactgatt	tgttgtaatt	aatacaactg	gtataaaaaag	79680
actgtgacga	ggccggggcat	ggtggctccc	gcctataatc	ccagcacttt	gggaggctga	79740
ggcaggcgga	tcacctgagg	tcaggagtct	aagaccagcc	tgaccaacat	ggtgaaaccc	79800
cactcttact	aaaaatacaa	aattagccgg	tcgtgtgggt	gcatgcctgt	aatcccagct	79860
cttcggggagg	ctgtggcagg	agaatcactt	gaaccgggga	ggtggagggtt	gcagtgagcc	79920
gatatcgccc	cattggcactc	cagcctgggc	aacaagagcg	aaactccgtc	taaaaaaaaaa	79980
aaagaaaaaa	aacacataaa	acaaaacaac	actgtgacgg	ttcccaaaaa	ttaggagcat	80040
aattaaaggga	actcctgata	aaaatttaatt	ttatcttaca	tgtaaaactaa	aatgacttta	80100
tgaagttaat	tcagaaatac	aatgcagggt	attagtttgc	cacagctgcg	tattcagcct	80160
aatgtaatat	tcttgtttatt	tttaaatctt	tcttttaact	ttactcatat	gtggatcatc	80220
aaatttcuaa	agattaaatg	acaatactct	tagcagcaag	cttccctaag	catataaaca	80280
tttttaattgg	tgatgattca	gaaggtagcc	gaagaatatg	tactgccaga	tatcattcac	80340
ccccatatac	ctgcccagaca	gacatcccat	tttgggaccc	tggataaatg	tgtgggtgga	80400
gagaaagata	ggagaaagtg	gtataagcaa	atggccttgg	agtctgattg	acagcgattg	80460
aaatcctgtc	tctacctctt	aacagcctca	tgatcctaca	taagttaacc	cgatcctcag	80520
ggccacatct	gtaaattggg	ggttgcgatg	gcagccatct	cacagggtct	cttttcgggg	80580
aagggcagga	attatggatt	aagttagcta	gtaattgtaa	agcacttaat	acaaggaggg	80640
cgcataataa	gtacttcata	aataatgacg	gccattatca	tgactgaggt	gtatgcagct	80700
gtcgggggatt	acggcgactt	cagaatttct	ggtgggcagg	gctcaaaggc	agcaaatcac	80760
actggaagtc	gaggtgaggc	actgcttctg	cacagactgc	ttagctggag	agaatgagga	80820
aggcttagag	gagatttaga	ggaacttaga	gtcctccgcc	tccaactctg	tgggatctgc	80880
tcccgtgcca	gagacattca	ggggatttct	cgcactctcc	cctcccctac	gtccctcccg	80940
ccccatccaa	ctaaccacac	aacacatata	aaatagcccc	tgcgaggttc	tgcacgctgg	81000
aagggaaacag	gagaaggggcg	ctgcgctttc	ttgctgatgc	cctgtacttg	ggcccctggg	81060
agacacagcc	acttgtcccc	tcagcctgca	gagaaatccc	acgtagaccg	cgcccgggtc	81120
cttggcttca	cccaatctcc	ctttgggtgg	ggtgggatgc	acgatccaag	gtttttattgg	81180
ctacagacag	cgggggtgtgg	tccgccaaga	acacagattg	gctcccagag	gcatctcgga	81240
tccctgggtgg	ggcgccgctc	agcctcccg	tgcaggcccc	gccgaggcca	ggaggaagcg	81300
gccagaccgc	gtccattcgg	cgccagctca	ctccggacgt	ccggagcctc	tgccagcgt	81360
gcttccgtcc	agtgcgctg	gacgcgctgt	ccttaactgg	agaaaggctt	caccttgaaa	81420
tccaggcttc	atccctagtt	agcgtgtgac	cttgagcagt	tgactttatt	tttcagtgcc	81480

39/122

tagttttcca	gataccagga	ctgactccaa	ggactattac	tcattctggag	ggtttagcac	81540
agtaccgtcg	catagtaaat	ttccatgtca	gttttgggta	cctttcatgc	acttgcaaac	81600
atgccatgct	ctgaaacgaa	atagggacat	cttttttttt	ttttttttta	aggagtcttc	81660
ctctcgccca	ggctggagtg	cagtggcgcg	atcttgggctc	actgcaacct	ccacctcccg	81720
tgttcgagat	tctcctgcct	cagcctcctg	attagctggg	actacaggca	tgccacgacg	81780
cccagttaat	ttttgtattt	ttagtagaga	cgggggttcg	ccatcttggc	caggctgggc	81840
taactcctga	cctcaggtga	tctgactgcc	tcagcctctc	aaagtgttgg	gattacaggc	81900
ataagccact	gcatctggcc	agaaatgaaa	taagtaaate	ttttaacctg	ctctaacaat	81960
atagtgaana	gaccatatta	ttattagagc	agggttaagg	atttgcctat	ttcgggttct	82020
agttatagtc	ttaaacttgg	acattcttgt	agaaagtaaa	aagtttcttc	ttcaaagttc	82080
cccttcttgt	taaaagaatac	atcataagtg	ttagaagtaa	tagttttattt	taaagactaa	82140
ctttcttcaa	gcctccttgc	tttgtgctaa	taactctttg	ttaagcccta	tcctatgtaa	82200
ctgttgggaca	tgctcacagg	cacgttccag	ttcacagcct	atgccccttc	cttatttgga	82260
aatgttattg	cttccttaaa	cctttcggta	agcaacttcc	tctccttctt	cgttcttctt	82320
tgcacttacc	tatttagaaa	gttttaggct	attagcaaat	cggctatcag	tttaagagtg	82380
tgaggctccg	ctccagccaa	tggatgcagg	acatagcagt	gaggacgacc	caaatgcgta	82440
agggataaat	atgtttgctt	ttcctttgtt	cagggtgtgct	ctcgacatcg	ttccatctgc	82500
gattgagcac	cctttctgca	gaaagtaaa	attgccttgc	tggagatctt	ttgtctccgt	82560
gctgactttt	cttcgtggca	ccgattatct	atttctaaca	attttggtat	ttctaacatt	82620
ctgaacaatc	ttgggctagt	tgctcttctt	gggctgtgtt	ccccatccgt	cacatgataa	82680
acttcattgg	tttaaaaacc	ccagcgaaca	tttattgagt	tactattacc	ttcctgcctt	82740
ccccaacccc	aaccccaggg	agcagttaca	acctcagccg	ctgagcgcac	tcgccgggtg	82800
ttaagaagca	ccaaagacag	ggaggcttga	ttgattttgc	tttgggagta	gagggtcaga	82860
agattcacag	gaaaatggca	tttgagcaag	gatgattcac	tggagctagc	ttttaaatat	82920
tggcgaggct	tttatgttgc	agtccttacc	aaagttagag	attcgagagg	actgcactcc	82980
gaaataagcc	cgcttcccct	tttcattcgc	taatgatcca	gggagctgct	ggttccgcat	83040
gcggcaggtt	gtgccttttc	ctaatacagg	ttctgcatcg	cctcgaaacc	gcaggccgtg	83100
gcgggttctc	ctgaggaagc	agggactggg	gtgcaggggtg	aagctgctcg	tgccggccag	83160
cgctgtgag	caaaactcaa	acggaggagc	aggaggggtc	gagctggagc	gtggcagggt	83220
tgacctgccc	ttttagaagg	gcacaatttg	aagggtaccc	aggggcccga	agccggggac	83280
ctaaggcccc	ccccgttcca	gctgctggga	gggctcccgc	cccaggagg	tagttttgca	83340
gagactgggt	ctgcagcgct	ccaccggggg	ccggcgacag	acgccacaaa	acagctgcag	83400
gaacggtggc	tcgctccagg	caccagggg	ccgggaaaga	ggcgcgggta	gcacgcgcgg	83460
gtcacgtggg	cgatgcgggc	gtgcgcccc	gcacccgcgg	gagggggatg	gggaaaagg	83520
gcggggccgg	cgcttgacct	cccgtgaagc	ctagcgcggg	gaaggaccgg	aactccgggc	83580
ggggcggttg	ttgataatat	ggcggttggg	gctgcctggg	catcccagg	aggcggtggg	83640
gcccactccc	gggaagagg	ttccttttgc	cgctagtcca	gcggccccct	tggaccggga	83700
agtcggggcc	tgagtgagg	tgaggggagc	cgggccccct	ccgcgcccag	cccccgcac	83760
cctccgtccc	gacccggggc	ccgccatgtc	cttcttccgg	cggaaaaggta	gctgaggggg	83820
cgccggcggg	gagtcaggcc	gggcctcagg	ggcgcggggtg	gggcagggtg	gcctgcgagg	83880
gctttcccca	aggcggcagc	aaggccttca	gcgagcctcg	acctcggcgc	agatgcccc	83940
tgagtgcctt	gctctgctcc	gggactcttc	tgggagggag	aagggtggcct	tcttgccgca	84000
ggtcagagga	gtattgtcgc	gctggttcag	aagcgattgc	ttaaagcccat	agaagtctct	84060
gcctgtttgg	ttagaacag	ttcttaggtg	ggggttagtt	tttttgtgtt	tctttgagga	84120
ccgtggatca	agatcaagga	aatctcttta	gaaccttatt	atggaagtct	gaagtttcca	84180
aatgttgagg	gttttatgtc	taaaagcaac	acgtgaaaaa	attgttttct	tcacccagtg	84240
ctgtcttcca	atttctctct	tgggggggagg	ggtagttaact	gctgttacta	aaataaaaatt	84300
acttattgct	aaagtctccc	aacaggaaga	ccactacttt	tgatgacttt	ggcaagtttg	84360
ctaactactg	gaaccctaac	ttacaaacga	actacttaca	tttttgattt	ccagttgtat	84420
tacctgcccc	atgtttacgt	agaaacagct	taattttgat	tctgggtaac	gttggtgcac	84480
ttcattaaaa	atacatatcc	gaagtgcaca	agtatgggtc	tgtggacagc	agtgattttt	84540
cctgtcaatt	cctgttgctt	cagataaaa	gtaccagaca	gaggccgggc	gcgggtggctc	84600
acgcctgtaa	tcccagcact	ttgggaggct	tggcggggtg	atcacctgag	atcgggagtt	84660
caagaccagc	ctgaccaaca	tggagaaaacc	ccgtgtctac	taaaaataca	aaattagcca	84720
gggtggtggc	gcatgcctgt	aatgccagct	acttggggagg	ctgaagcagg	agaatcgctt	84780
gaacctggga	gcgggaggtt	gcgggtgagc	gagatagcac	cattgcactc	cagcctgggc	84840
aaaaagagcg	aaactccgtc	tcaaaaaaaa	agtaccagac	agaaatgggt	tttgttttct	84900
ttttttgttt	tgagacggag	tttcgctctt	gttgcccagg	ctcgagtcca	atggcgcgat	84960
ctcagctctg	gctcactgca	acctctgtct	cccagggtta	atcgattctc	ctgcctcagc	85020
ctcccaagta	gctggggatta	cccatgcccc	accatgcccc	gctaattttt	gtattttttag	85080

40/122

tagaacggg	gcttcacat	gttaggctg	tcttgaacc	ctgacctca	gtgggctcc	85140
cacctcgcc	tcccaaagt	ccaggattac	aggcatgag	caccgcgcc	agccagaaat	85200
gggttttga	aaaagcacta	aacaaaatcg	aacttggtt	catatgacag	ctctgctgct	85260
aactgtaaca	ggggcagacc	agttaaccta	cttttctgtc	ttctgtcagc	tgagaattag	85320
atgattccca	aaggccccatt	gaactctgaa	tgactttaaa	tactttctct	taagtgggta	85380
cacggttttg	gtaactgatg	ccagggtgatg	aatgcatgaa	agtgtttaat	gaatgaaacc	85440
ggtaaaatag	taggaggaag	ctttatttgg	aaggcagggg	tatacctaata	agctctctaa	85500
tttatttggt	ttgaagtgg	taacttttgt	ttttttaagg	ggggaaaaca	ttctaagaat	85560
aatgaggcaa	actgcatatt	gcacaagaga	ctgttgtctc	tattcaacaa	ataccttttg	85620
agtgtccaga	gtctgccagg	tgctgtgcta	ggccctcacg	attgagtagt	gaaccagaga	85680
atgtccctgc	acccatggag	cttattgtct	actggggtag	acagataata	aataagcaaa	85740
caaactctct	ctcttctccc	tttcgctcca	tgtaagtgtg	tgtgtatagg	tgtatactta	85800
caagttgagt	aaagtgttat	gaaagattaa	gaggagaaat	gcattttggg	tagatgttag	85860
aggactcagc	agggtgacct	gaaacttaga	gctgaaggat	cagtaggagg	taactagaga	85920
ggccagggaa	tcgcatgttc	aaaggccagg	aggcaagaaa	gagcatgggt	cccttcaaga	85980
gaggaagaa	ggctactgtg	actggagcat	agatgtaggc	aagtgttggg	tgattgagag	86040
ctctacgggc	catgggttag	ttttattcct	aatgccgaga	tgccaaacat	gggtggtcat	86100
atctgtaatc	ccagtatttt	aggaggccga	ggcaggaata	tagcttgaac	ccaggagttc	86160
aagaccagcc	tgagcaacat	gagacctgta	caaaacattt	aaaaaattgc	tgggtatgat	86220
ggtgcacacc	tgtgggtccc	gctactcagg	aggctgaggc	agaaggatca	cttgagccta	86280
ggaggtggag	gctacaatga	gccatatttg	agtcactaca	ctccagcctg	gatgacaaag	86340
tgagaccatg	tgtcaaacaa	aatacagaaa	gaatattaat	ttaaaatttt	gaaagaggag	86400
tgatctgaac	ttatatctta	aaaagatcat	tctagggcat	gggtggctcat	gcctgtaatc	86460
aagggtcttg	tgaggctgag	acaggaggat	cacttgaggc	cagttcgaga	tcaacctgta	86520
cagcatagag	agactccatc	tctacaaaaa	gaaaaataaa	atagctgggt	gttggtgatt	86580
attcaggagg	ctgaagcaga	aagatcactt	gagcccagga	gtttgaggct	gcagtaagct	86640
atgatcccac	cactgcaaca	cagtggatgc	ttgtctcaaa	aaaaaaaaaa	aatcattcta	86700
ggtgcttttt	ggaggctgga	tgtggtaaga	gtagaagctg	gagatgggtc	tgtagggat	86760
tcgattcaga	ctttaaatat	catcaatgca	ttgagtccca	aattttacatc	actacgttgg	86820
atccttgccc	ctgaatccag	actggatat	ccaactttag	gttcagtttg	tatctctacc	86880
tgaccaatat	agagggtgtcc	agtcttttgg	cttccctagg	ccacattgga	agaagaattg	86940
tcttgagcca	cacatagagt	acactaacgc	taacaatagc	agatgagcta	aaaaaaaaatc	87000
gcaaaactta	taagtgttta	agaaagttta	cgaattttgt	ttgggcacat	tcagagccta	87060
cctgggcccgc	gggatggaca	agcttaatcc	agtagatacc	ttcaacttac	aatatctaaa	87120
attttatgcc	agatttagtc	attttaaac	tgctcatcag	tttttctcaa	gaagtagtat	87180
tttggtcttt	tttcttttct	tttttttgag	atggagtttc	gctcttatcg	ttcaagctgg	87240
agtgcatgg	cggatcttgg	ctcactgcaa	cctccgcctc	ctgggttcaa	gtgattctcc	87300
tgcttcagcc	tcgcaagttag	ctggaattac	aggcatgcgc	caccatgacc	agctaatttt	87360
tgagacagg	gtttcaccat	gttggtcagg	ctgggtttgt	actcctgacc	tcagggtgatc	87420
tgctgcctc	ggcctcccaa	aggctgggat	tacaggcatg	agccaccgct	cccggctgca	87480
tttttggtt	tttagttgct	cagcccaaaa	cttttagtaca	tctttgaacc	tcttctttcc	87540
tcctactcta	tatctgatcc	atcagcaaat	ctgttaggtc	tacctcacac	atatcgaaat	87600
cctaccacgt	ctcaccatct	gtgacaatta	acaccctggg	ctaggcagtc	atctctgtta	87660
agattgagtg	gttaaggatg	tcctctaagg	agatgacatt	caaactcttag	cttaaagtgc	87720
aagagggagc	tggttttata	aagattgagg	aggcagcatt	attttgccat	aggcttccat	87780
ttggtttcca	ttccattctt	gatacttatg	gtatatattc	aaaacaaatg	cacagaaaca	87840
gacccaggta	tattgggaat	ttcggtatata	gagttcctag	ttgggaaaaag	atagactgat	87900
ctgtaaatga	tgctagtatt	ccatcatctg	gcaaaaaata	atctcctgcc	tcctctcata	87960
tatctcagat	caacagactt	tttctgttaa	gggccaaatc	ataaatatct	taggctttcc	88020
agaccatag	gtttctgtca	cactctcctt	tactcctgaa	gccatagaca	atatgtaaac	88080
aaatgggcat	ggctgtgcta	cgataaaact	ttacttacaa	aaactggtag	tgggccagtt	88140
taggcatggc	cagcactttg	ggaggctaag	gcagatggat	cacttggggg	caggagtttg	88200
agaccagcct	ggccaacatg	gtgaaaccct	gtctctacta	aaaatacaaa	aaatagctgg	88260
gcatgggtgg	gggtgtctat	aattccagct	actctggagg	ctaagacaca	agaatcactt	88320
gaaccagga	ggcagagggt	gcagtgaagt	gagatagcac	cactgcactc	cagccagggt	88380
gacggagtct	taaagcaaaa	caaaacaaaa	ggtagtgggt	tgtatttggc	ccatgggctg	88440
tagtttgcca	atccctgatg	cagaaacaaa	ttccaggtaa	ataagagcct	ggaatgttaa	88500
aaaaacaaaa	cttgaagtca	tgtagaagaa	caggtagggg	gaacaatcct	gatctcagga	88560
taggaaggga	tattgcttaa	aataagacac	aggaaaatat	aatccatgtt	gtgtaaat	88620
gactacgtta	aaacttaaaa	ctttcgccaa	gcgcgggtgg	tcacgcctgt	aataccagta	88680

41/122

ctttgggagg	ccgaggtgag	cagatcacca	ggtcaggaga	ttgagacat	cctggctaac	88740
acgggtgaaac	cccgtctcta	ctaaaaatac	aaaacattag	ccgggctg	tggcgggagc	88800
ctgtagtccc	agctacttgg	gaggctgagg	caggagaatg	gcctgaaccc	gggaggcgaa	88860
gcttgcaagt	agctgagatc	gcgccactgc	actccagcct	gggcgacaga	gtgagattcc	88920
gtctcaaaaa	aacaaaacaa	aacaaagcaa	aaaacctaaa	actttcatac	aataaagtat	88980
acctaagata	cttctagaag	agaagattta	catccaggac	gtgtatggaa	tttctgcaag	89040
taataagtaa	aagacaaggg	acatgaagag	gcagttcaca	aaagaggaag	ccaaaatgac	89100
caataaacat	gaaaaggatgt	ttaacctcaa	aggaaacaag	gaaatgaatt	aaaaacatca	89160
aatgccattt	caaaaactagt	aagttggcaa	aattaaaaat	accaaggatg	agaatatgaa	89220
gcattggctat	atgagtgcac	ggaatggtag	agtcactttc	attaaaaatg	cacataattt	89280
gtttttttatt	tattttttttg	agacagtcta	tgtcgccag	gctagaatgc	agtggcatga	89340
tctcggtcca	ccacaatctc	tgcctcctgg	gttcaagcaa	ttctcctgcc	tcagcctcct	89400
gagtagctgg	gattacaggc	acatgccaca	acgcccgggt	aagttttgta	tttttagtag	89460
agucagggtt	tgtccatggt	ggccaggctg	gtctcgaact	cctgacctca	ggtgagctgc	89520
ttcccaaggt	gctgggattta	gaggcgtgag	ccaatgctcc	tggctgaaaa	aaatgcacat	89580
aatctgttac	ctagcaattc	catgtctaga	ggcttatcct	agagaaattc	ttgcttatat	89640
gcataagtag	acgtgtacta	gaatgttcac	tagttgaatg	tttaagtga	aattaggaaa	89700
taaaagtaaa	gttcattaac	aggaaaatga	gtaaaggat	atttataaaa	caattaaagta	89760
gctaaatga	ataaactaga	gctgcgtgaa	tgaactagaa	ctggttcaat	agtcatgtca	89820
guttattgaa	tgaatacaga	tcagatatgt	attagatgtc	atttgtgtaa	ttaatttttt	89880
tttttttttt	gagatggagt	ctcactctgt	tgccagggt	ggagtgcagt	ggcgtgatct	89940
cagctcactg	caacctccac	ctcctgggtt	aaagtgatcc	tcctgcctca	gcctcccag	90000
tagttgggtt	tacaggcatg	caccaccatg	cccagctcat	tttcttattt	ttagtggcca	90060
cagggtttca	ccatggtggc	caggctggct	ttgaactcct	gacctcaagt	gttccacca	90120
acttggcttc	ccaaagtgtc	aggattacag	gcgtgagcca	ccgtgctcag	ccatttgcgt	90180
gattttttta	gatgtgcaga	ataatgccat	taaaaaaaat	acacatacat	gtatatatat	90240
acacgttttg	ctgggtgtgt	tggctcacac	ctgtaatccc	agcacttttg	gaggctgagg	90300
caggaggatc	acttgagccc	agggtatcaa	gactagcctg	ggcgagatag	caagacccca	90360
tctcaacac	agaaaggata	attaggtatg	gtggcatgag	aggatcactt	gagcccagga	90420
gttcgagtgt	tatcaggcca	ctgcactcta	gcctggacaa	caaagcaaga	ccgtgtctca	90480
aaaaaataaa	aataaaaaagt	atttgtatgt	ggtcatagtc	aaaaaacgta	catggaagga	90540
aaatgtcttt	atttattttt	ttattttttt	ttttttaaga	cagagtcttg	ctctgtcacc	90600
caggctgggg	tacagtgggt	taatctcagc	tcaccgcaat	ctcgccctcc	cgggtccaag	90660
cgattcttct	gcctcagcct	tctaagtagc	tgggactaca	ggtaccgcc	accacaccct	90720
gctaattctt	gtgttttcag	tagagacagg	gtttcaccat	gttggcaagg	ctggtctcga	90780
actcctgacc	ttaagtgaag	caccgcctt	ggcctcccaa	agtcctggga	ttacagggtg	90840
gagccactgc	gcttggccag	gaaatatcta	atttgatgag	tatttatatc	tgggaaagga	90900
agggctcagg	ggtgattcat	aggaactcta	aagtctatgt	ataatactta	gggggacaga	90960
aggaataaaa	gcaaaatgct	gatatttgat	tgttgagttg	tgtatatgtt	agaagtataa	91020
cataggagat	ctgattgata	gtaggagaat	gttttttaggt	ggtaaaagtg	gaaccgtggt	91080
ggtttggttt	ggcagtagaa	tcagttgggt	atagtttgta	tgtggaaggt	aataaacaga	91140
ccatgttaag	gatgacttcc	ggaatttttg	tctgagtagt	gggtggatga	cagtgtcatt	91200
catgagggaa	gatgaagact	gaggtaggaa	caggtttggg	agaagatgac	atgttccctt	91260
ttagacaagt	ggaattatgg	aagatggcag	gtaggtggtt	agctatatga	atttgagata	91320
aaagatttag	gatggagata	taaatttagg	agtaacagcg	tatctatggt	attgtaagcc	91380
ttaagaatgg	gtaggatcag	ccaggaaata	cagatgtata	tgcagaagag	aggagtcaag	91440
gaagccaaga	caagttaatg	tttaaagtga	gtgatgtagt	ccatgggcag	atgctgctga	91500
gagggctgca	aacaccagtg	accctacaac	atttttaaat	gtcgtcttcc	tgacagcagt	91560
gatcagtacc	tgcaacgatc	ttattttatt	ttttcatggt	agtctccaca	cacttgaatg	91620
tagacttttc	gaaggcaaaa	tcattgcctt	ttctgagctg	ggagcatgtc	tggcacatac	91680
caagcactca	acagttagtg	tattgacttc	atccagatac	tctgagggag	agttattttc	91740
tgctactagc	ctttcacctt	tcaatgttta	agagcacaaa	tacagagatg	ggcacgtttt	91800
ggcatttctt	attttgataa	ccttttcctg	gtaagatttt	ttaatgttga	aaaaaaaaaa	91860
caagaaaaga	gggttaaaaa	tagtcttatg	tcagatcctg	tgatagaatt	cacacttggc	91920
ttaaagctgt	gggcaccttc	ctatcttgga	tgtcatatta	gcttatctac	agcagaattt	91980
ttactgtttt	atgtagttaag	gaagcaatta	tatgattatt	ttacagacaa	attattcttt	92040
atctttttatt	tttttagacg	gagtctctct	ttgtctccca	ggctggagta	cagtgtcgcg	92100
atctcggttc	actgcaacct	ccgcctcctg	ggttcaagca	attctctgcc	tcagcctccc	92160
aagtagctgg	gcttacagg	gtccgccacc	acaccagct	cattgttttg	tatttttagt	92220
agagatgggg	tttcaccatg	ttggccaggc	tggctctgag	ctactgacct	caggtgatcc	92280

42/122

accgccttg	gcatcccaa	gtgctggaat	tacaggcgtg	agccaccgtg	cctggcccg	92340
acaaattatt	atactctgag	tgtagaggc	ttaggatgtt	ttcacttgat	gctatgggag	92400
gaataagtaa	taagatatga	tacacaacca	aagacctttc	ttcactatgc	ttctagtagc	92460
tagtactatg	gatgacacat	ggtaataata	ttggtagca	tttgtcctca	atctactgtg	92520
ctagtactc	ttctaagccc	cttacaggta	tatatTTTT	ttcatcaata	atcctctaag	92580
gtagttttta	ttattgacct	aattttataa	atcaagaaaa	ttaagaccca	gagaagtaag	92640
taacttgctc	aagatcacat	ggcttataag	tggtagagcc	agaatttgac	cccagatgtt	92700
gtgactacat	tgtctctcca	taagcagggt	caactctttt	gactggatgc	tggtccaagg	92760
tcacttcctt	agagaagcct	ttgctgacaa	ctaccctcct	gtgccctcct	ccaaggctgt	92820
ccattgttct	agaactttga	atactcatct	tagaataaag	ctggctaat	ttttacagtg	92880
ttatagaatg	gatctctgac	tgcaaaagtt	ggtcataatt	atctttttat	gttctagtga	92940
aaggcaaaag	acaagagaag	acctcagatg	tgaagtccat	taaaggtaag	ttctgccctt	93000
ggcagtcac	tgcattaaaa	agtgatgtgc	tttgcatttg	tgagttcctt	aatcctgtta	93060
tactctctct	tttggcattg	atcatttctg	ccttatTTTA	taattactta	tgattttgat	93120
ttatttccct	ctttaacctg	tataatgctt	taactcttag	catataataa	gtaggctttt	93180
tttttttttt	tttttttgga	gacggagtct	tgctctgtta	cccaggctgg	agtgcagtg	93240
cgcgatcttg	gctcactgca	agctctgtct	cccgggttca	caccattctc	ctgcctcagc	93300
ctccccagca	gctgggacta	cagggtgcacg	gcgccacgcc	tggttaattt	tttgtatttt	93360
ttagttagga	cagagtttca	ccatgttagc	cagtatggtc	tcgatctcct	gaccttgatga	93420
tcggcccgcc	tcggcctccc	aaagtgcctg	gattacaagg	gtgagccacc	gcacccggcc	93480
gtaagttagc	tttttttacc	ttaatTTTT	ttttttgaga	tgaggtcttg	ctcttatccc	93540
caggctggag	tgcatgggtg	ccatctcggc	tcactgcagc	atccacctcc	cgggttcaag	93600
cgattctcct	gcctcagcct	cccagtagc	tggtgattaca	ggtggccgcc	accatgcccc	93660
gctaallttt	gtatttttag	tagagacagg	gtttcacctg	gttggccagg	ccagtctcaa	93720
actcctgacc	tcaagtgatc	cactcgcctt	ggcctcccaa	agtcctggga	ttacaggcgt	93780
gagccaccat	gcctggccat	aagttaggctt	ttactgagcc	ttgtgtgtat	tggtatcct	93840
agtgtattaca	gtgaaccagt	gcccttctta	ttaatcacac	atTTAattgt	tccttaaaag	93900
tgattagttc	actttatttta	tttagtaaga	caaaaaatga	agaataactc	taactgagca	93960
gtctgttaac	tgtaggaaag	cactgacact	tataaggctt	agttttctgt	catttatcca	94020
gaagtatgg	tgattacagt	ttttactttt	ttatttgaat	gaacaacctt	aatttaaaat	94080
atatatttgt	tattttttgt	tggtatcgat	acattgtcct	tgtttataga	ttagagcatg	94140
ctttttaaag	atgtctgatt	actcactgat	ttattttgtc	cagtgtacag	agattgaagt	94200
gggaaaaatta	taattggaaat	tgtttccata	gtcattacat	attaatttca	tcaatttatt	94260
tcataaaaaat	ctgtagattg	ctacttattt	agatttttcc	ttcaaagtgt	tttatgttgt	94320
attgcttgca	ctgagtattt	attctatatg	ctcaatttgc	tgagagaaga	gactaattat	94380
aacttaggca	agttgtaaaa	ttagggaaaa	aagtaaggta	ccttacagcc	tagtttactt	94440
atctcttatg	taaagccagt	tagattccac	attagttcaa	actgccttct	ttgagcaaaa	94500
cttgattggc	agtataaaag	gcttaaagcc	cttctcaagc	agagacctgt	aaagactaga	94560
tctgactgta	gtagaaggaa	ggaacttaga	tgtttcaggc	agtgagaaca	ccagtcttcc	94620
actctaaact	ttgccactaa	cagtatgacc	ttgggaagtt	gtaactttct	tcagattcct	94680
catttggtga	ttggggggat	tgtctctaaa	aatttctaaa	ttctacttgg	gctaaaaaat	94740
tctgtgctta	tactctgatt	atgaagtaca	taatctgtgc	ttaacattca	ctgacttatc	94800
cttaggataa	tacagaagca	gtacaagaaa	cagccctca	agatgtttgc	agtctgggtta	94860
gaaagacaaa	cttatacaca	gaacagtatg	aaatagacca	aaataataat	agctgccatt	94920
tatagaacac	ttcttctgtt	ctgggcatta	gacaaaaact	gactataacg	gtgaacaaaa	94980
aagacttagg	tcctgccctc	attgaactta	cagattagta	ggggagagga	acattaatca	95040
agtaattcca	cagatggctt	agcctagatt	ggtagtgatg	gaagttaaaga	gatgtgaacg	95100
gacttgaaaa	aaaattcgga	ggcaaaatgg	atagaagttt	attattgatt	aaatatgagg	95160
tgtgagagag	agggatattt	aagattgata	cctaccttct	ggcttgccct	acagaaccaa	95220
aacaggaaat	tatatgttca	gttttgttat	gttgggtggg	aggtgctttt	gagtcattca	95280
tttatatatg	ttatatatgt	tattttatat	gcataagta	tttaaggctc	gagtttttaa	95340
ccaaagggtta	gagagtgatt	tttttagagtc	tagcaaacct	aagttgaaat	cctgcctgtt	95400
gaaatggctg	tttactagct	cattaacctta	gggcaaaagta	ttcaacttgt	tttcattttt	95460
gtcttcatct	ctaaaatgag	gaaaaatatg	tcttacaaga	ttgtcctgag	agatagatga	95520
aataatatcc	aaaaaaaaaa	aagggtacata	gagaaactcg	tatagtgcct	ggtatatagt	95580
aggctctcca	ttggtagcta	tcattatcta	gttttaacat	agccttcagt	ttgttgaatt	95640
agtcaaaactg	agtgaagcac	tgcaagggaat	tcagaggaat	ttgagatcaa	caaagtattt	95700
ctgaagttaa	gggaagactt	catggcaatg	acacttacct	tgtataaaaag	ttgaagaata	95760
agaaagattt	gaatgagaga	ttctttctct	tctccctacc	agcccgactt	cttatttgag	95820
gatatatattg	gcaaaaggggc	cttcagacaa	gtagaggggag	atTTTtacag	aaagattgag	95880

43/122

atgaagggtat	agaaggctgt	aaagaccaga	aaagagaatt	gagacagagg	aagcaggaag	95940
ccactgtagg	tttttgagca	agatattgat	gctgtaagta	tgggtgtttat	gaaagggttag	96000
tctggaagag	atttgcagga	tggagacccc	ggaagttttt	ttgttataat	acagaaagac	96060
ttgcaactgag	ggtagagggt	taaaaataaa	caggtaagta	aatgttttaa	catcttgaag	96120
gaaaagtcaa	caaactcttg	caagtaaaca	gataacagtg	aaaaagaatg	ggaccaagat	96180
tttgagtttt	ggagactggg	ggattgaaca	gacagggaaa	tgagaggag	aatcagatga	96240
tgatgtttta	agttgatatt	tagacagatt	gtgcttgaga	tggtaaagtc	aatgtgggtg	96300
ggaatgctta	gtagcgagta	atcagtgata	caagacccaa	gcccagggtca	aagacaagtc	96360
acagatacag	atcagggtct	tttcatctgc	tccacagagg	tgtaccctag	gagctgttgc	96420
aaacagtcca	tgtggagggt	gtgagtaaga	tgtttccctt	gaatttgcca	gaattacttt	96480
tttgttgttg	ttgttgtttt	ttctgagaca	aattctcgct	ctgttgccca	ggctggaggg	96540
cagtggcgag	atcgcgagc	tactgcaac	ctctgcctct	cgggttcgag	tgtattctct	96600
gcctcagcct	cccaagtagc	tgggattaca	ggcttgtgct	accaagccca	gctaatttct	96660
tttgtatttt	tagtagagat	ggggtttcac	catgttggtc	agactgggtc	cgaactcctg	96720
gcctcgtgat	ctgcctgcct	cagcctccaa	agttctggg	attacaggcg	tgaaccactg	96780
caccgggtcc	cttggttaagt	ttatttttgt	gggaagcaaa	ggaggtttca	gcttttataa	96840
agtttgaaaa	ttattgctct	ggtaataatt	aaagatttga	gagtaaatat	gctttcttagc	96900
agaaagaata	aaagaagaac	agatagcctc	aagaagggga	gccaagaag	caggctatat	96960
ctgacacact	gggtgttgat	aaatgggtat	taaaagaatg	agagcaatga	gcagatagaa	97020
gagaaatta	ggagagtata	ataccatgga	gaccaagaaa	gatagactat	caggaaggag	97080
tggtaaaaat	aagttactag	ttctaagaga	gatgttaaga	gggaccgggg	aaagccttgt	97140
acaaatgagt	tagtagcatt	ttacattata	tacatctaata	taagaaacaa	tgcgagagtc	97200
tcaccattcc	tatagactct	tacttgtact	tgtctgaaca	cgaaaactgg	cttttgttta	97260
taaataagct	aaaaattatt	ttgtccaat	ttctcatgaa	aataaaaaata	aaccttcttt	97320
taacattgaa	aaaatagttt	gaagacagtc	actcttcatt	ttgtaattcc	cacaactatt	97380
attgaatgac	tgaattatc	tttattctga	agccaaaggg	gtgatactga	tatttcttca	97440
gactactaaa	aatatatttt	atgaattttt	agtgtgcttt	atcttttttt	gttttttttt	97500
tttgatgga	gtttcactcc	cgttgctcag	gctggagggc	agtgggtgcaa	tctcagctca	97560
ctgcaacctt	cgcctcccag	attcaagcaa	ttctcctgcc	tcgggtctccc	aagtagctgg	97620
gattacaggc	acctgcccc	acacccagct	aattttttgt	atttttagta	gagacagggg	97680
ttcaccatgt	tggtcaggct	ggtcttgaac	tcctgacctc	aggtgatcca	cccaccttgg	97740
cctcccaaag	tactgcgatt	gcaggcatga	gccaccatgc	ctggcctgag	gaatattttt	97800
ctagggtccc	cccaccccaa	gcattttatt	tgcaatttta	gttttgttcc	taaagcaagc	97860
aagggttaag	gatttaaaaa	taatccgtat	tttagaatgc	tttctggctt	tgttactttt	97920
tatccacagt	agaagttctc	agagaatgat	ctccctcttt	taatttaact	ttttggcaca	97980
glattttgag	aattataaat	aatattagaa	tgttttctgg	ctgggtgtgg	tggctcatgc	98040
ctgtaactct	ggctacttgg	gaggctgagg	caggagaatc	acttgaacat	gggaggcaga	98100
gggtgcagtg	agccgagggtc	atgccactgc	actccagcct	gggtgacaga	gcaagactct	98160
gtctgggaaa	aaaaaaaaaa	aaaaaaagag	tgttttcttt	cctatttttc	accacttgat	98220
taagttactt	ttcctcttaa	gtattttttg	ctgagtatgc	tgacttaaga	gtaatgttac	98280
aaaatttaatt	ttttaaagtt	ctctgaaagc	ccctttatga	gagtttttagg	ctatcaaatt	98340
gtgtttaatt	cttaacaatt	ttttgaaaaa	tttatagctc	aatatccgta	cattccccac	98400
aaaaaagcac	taaaaatcat	gccttgctgg	aggctgcagg	accaagtcat	gttgcaatca	98460
atgccatttc	tgccaacatg	gactcctttt	caagtagcag	gacagccaca	cttaagaagc	98520
agccaagcca	catggaggcc	gctcattttg	gtgacctggg	taagtaacta	tcatttttta	98580
tttaacttga	ttagaaggat	ttgagtaaaa	tatgtgaaac	ttctgtcata	ggatacagaa	98640
ctatataaatt	ggaaagtgtc	ttggaaaaaa	tgtattttaa	ataacagcta	caagtataat	98700
gggtagctgt	gttgtgttcc	tgtaaatata	gaatataaag	catgcccagt	agaaaaacaa	98760
gcattttccag	aagaaatata	tctgatcact	aatataaat	atatgaaaaa	gatgtctcac	98820
tttatttact	aggggaagtgc	aaattaaaaa	aatcagttaa	tgttctccta	acacattagc	98880
atatttttta	aagtttgaca	atttgaatgt	cagtgaagat	gcagggaaat	acccctccta	98940
tttagtgata	atataatctg	gtgaagactc	tttggaagc	aatttggaaa	tcagtataaa	99000
atatgcatgt	catttaggcc	actctttcta	agacctagcc	ctcagatatg	ctcattcata	99060
tgtgcagggt	tgtatgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtatatgta	tgtatgtatg	99120
tatgtatgta	tgtatgttga	aggctattca	ttatagattt	gtttgtgata	gcaaaaaatt	99180
atggacaaca	tataaatatc	tgttataggg	aaataaccaa	attgtgggtat	acgcattgctc	99240
tggagtataa	tatagccatt	tgtttctatt	tatttatattt	cttgagacag	ggttttactc	99300
tgttgcccag	gctggagtgc	agtggatga	tcattggttca	ctgcagcctt	cacctcctgg	99360
gcacaagcca	ttctctcgcc	tcagcctcca	gagttactag	gactgcaggc	atgtgtcacc	99420
acaccagat	aattttttta	ttttttgtag	agacagggtc	tcactatggt	gcctaagctg	99480

44/122

gtctcaaact	cctggcctca	agcaattctc	ccacacaggc	ctcccaaagt	gctgggatta	99540
ccaacgtgaa	ccaccacacc	tggttcagtg	tagccattta	gaaatctaaa	aaagacgtgg	99600
gaaaatgtct	aaggcatgtt	taaattgtgag	aaaagcaagt	cacagtatgc	atggtaaaat	99660
ccgttatatt	aaaataagtt	cttccaaaac	aaaaacatat	gcaggagacc	tttattttgt	99720
cagtatttct	tacccaaatt	tctgcactta	gaaaattgca	tgtcatgttg	tcataagttg	99780
aaaaaaagat	ccatgaacca	atggacttct	aataaaaatca	gtcctgcttt	tgacatctct	99840
ctctactttt	gtgtatatct	aaaccagagt	gtcaatgtgt	ttgtggggca	cacttagcaa	99900
taatacatag	cagacaaaat	gcatatagct	cagagagtaa	aattgttaagt	tttgctagat	99960
cactcataaa	ttgctgatga	gaatttataaa	tgggtgcagat	gctctggaaa	acaggcagtt	100020
tctttctttc	tttttttttt	tctttttgag	acaggggtctc	actctgttgc	gcaggctgga	100080
gtacagtggc	gtgattacaa	ctcactgcag	cctcaccctc	ctcagggttca	ggtgatctct	100140
cctcagttct	ctgagtagct	gggactatag	gcatgcacca	ccacgcctgg	ctaatttttg	100200
tatttttttt	tttttttttt	gtagagacgg	gggttctgcca	tgtttcccag	gctggtctca	100260
aactcctgga	atcaagcgat	ccacttgctt	aggctctccc	aagtgtctggg	attacgggcg	100320
tgagctactg	tgcttgccct	aggcagtttg	ttgtttgttg	tggtttgttg	tttattttat	100380
tgtagacgga	gtctcacagg	ctggagtgca	gtggcccaat	ttttggctca	ctgcaacctc	100440
cgctcccag	gttcaagcta	ttctcctgcc	tcagcctcct	gagtagctgg	gatgacaggt	100500
gcctgccata	atgcctggct	gatttttgta	tatttagtag	atatgggggt	tcacatgtt	100560
ggctcaggctg	gttttgaact	cctgacctca	gggtgatcagc	ccgcctcggc	ctcccaaagt	100620
gctgggatta	caggcatgag	ccgtcatccc	tggtgtgtgg	tttcttatga	cgtgaaacat	100680
gcaattacca	tatgacctag	cagttgcact	ctgtatttat	cccagataaa	tgaaaactta	100740
ccttccaata	aaaacctgtg	cacaaatgtt	catagcagct	taatattgaa	aaactggatg	100800
ttcttcagca	gggtgaatgaa	ctggttcatt	cataccatgg	aataccattc	agcaataaaaa	100860
aggaacaaac	tggtgataca	tttaaccacc	tggatgaata	tcaagggaat	tatgctgtca	100920
gacaaaaaac	agtcctctaaa	gactacatat	agtatgattc	cgtttgata	atattcttga	100980
aatagagaaa	ttaagagaaa	tgaaaagatt	agtgtttgcc	agatgttaga	gacagggagg	101040
tgagaggggt	aagtgggtgt	agttataaaa	gtgcaacatg	agggatcttt	gtgatgttga	101100
agttgtatct	tggcagtgga	tgcaaaaatc	tcaatgtgat	aaaattacaa	agaactaaaa	101160
acaagaatga	gtatagataa	aactggggaa	atctgaacaa	gttagagtgt	tgtatcactg	101220
tcagtatctt	agagtgatat	tgtactatag	ctttgcaaga	tgttaccatg	ggagaaacta	101280
aagtgtacaa	gggatctcta	ggtattatta	tttttttaga	gatgggggtt	cactatgttc	101340
cccaggccgg	tcttgaactc	ctgggctcta	gtgatccgcc	tgccccagcc	tcctaaagta	101400
ctggaattac	aggcgtgagc	gaccatgcct	ggccctttca	gtattgtatc	ttagaacttc	101460
atgtgaatct	agcattatct	catagaattt	aattaaaaga	aattgttaaac	ctcacagaag	101520
atcagaattt	cctcaagttt	gtgatgttga	caaagatgaa	ctagttgaca	ctgacagtta	101580
gactgaggat	gaagacacga	cgtgcttcaa	aaaaatgatt	tgaatatcaa	tggattaaga	101640
agaactcttt	tgacaaattg	atgaaacctc	cagtcagttt	tataagaatg	cccactctta	101700
tgatcatgct	atgaaagcca	atttttataa	aaattttttg	tctttcctaa	caattagctt	101760
gtggttataa	tttaaattta	gttaaatata	agataaatga	ttttttatta	agtttagttt	101820
catttttcaa	ggtacgatct	caaagctact	ctttaaccta	ctatgaatga	ataatgctga	101880
gttcataaca	tattttgtaga	tatatccaca	attttccctc	aggataagtg	cctacaagtg	101940
gaattactgg	actgaaaata	atgcagtttg	ctaagacttt	gctatctggt	cctgaatgct	102000
cctccaaaaa	ggttttgcca	gtttacatcc	tcatgaccag	cgaatgagag	tgttgccctat	102060
tttctgtgct	ccttggtact	gcttaataat	ttttgaaaaa	aatctaattt	gacagacaaa	102120
aatgcatttt	atgttaattt	gcttttctgg	gatttttaat	gaggttgagt	atagttttta	102180
atatttttat	tggccctttt	ggaactagta	tcataagttt	tttttcttaa	gaatttatgt	102240
agtctgggct	gggcgcagtg	gctcacgcct	gcaatcccag	cactttggga	ggccgagggtg	102300
ggtggattgc	cgaaggctcag	gagtttgaga	ccatcctgac	caacatgggtg	aaaccgaatc	102360
tctactaaaa	gtacaaaaac	tagctcagcg	tgggtggcggg	tgctgttaat	cccagctact	102420
taggaggctg	agtcaagaga	atcgcttgaa	cccgaggagt	ggaggttggt	tgcattgagc	102480
cgagatcgcg	ccattgctct	ccagcctagg	caacaagagt	gaaaagtctc	aaaaaaaaaa	102540
aaaaaaaaaa	aaaaaagaat	ttacatggtc	tgaattgcca	ttaaaagaga	tatgagaatt	102600
attgagtaac	aaataacttt	ttaataattt	aggcaagttt	tggacgattg	tactttgttt	102660
agaaaccaaa	agcatagtat	ttgtagtgtt	tttattttact	ttagttgcta	ggaagtaaac	102720
tttattcaag	gtctctggta	ccagttgttg	ctaaaagtga	ttgactaatc	tgtcaatctg	102780
aaattatttg	ttgctgaact	gctaattctt	ttgcttctat	cttttaggca	gatcttgtct	102840
ggactaccag	actcaagaga	ccaaatcaag	cctttctaaag	acccttgaac	aagtcttgca	102900
gcgactattt	gtcctccctt	acttcattca	attcatggaa	cttcggcgaa	tggagcattt	102960
gggtgaaattt	tgggttagagg	tcgaaagttt	tcattcaaca	acttggtcgc	gaataagagc	103020
acacagtcta	aacacagtga	agcagagctc	actggctgag	cctgtctctc	catctaaaaa	103080

45/122

gcatgaaact	acagcgctctt	ttttaactga	ttctcttgat	aagagattgg	aggattctgg	103140
ctcagcacag	ttgtttatga	ctcattcaga	aggaattgac	ctgaataata	gaactaacag	103200
cactcagaat	cacttgctgc	tttcccagga	atgtgacagt	gcccattctc	tccgtcttga	103260
aatggccaga	gcaggaactc	accaagtttc	catggaaacc	caagaatctt	cctctacact	103320
tacagtagcc	agtagaaata	gtcccgcctc	tccactaaaa	gaattgtcag	gaaaactaat	103380
gaaaagtgag	tatgtgattt	tcttgtgtgt	acatatgtgt	ctcactttct	ttttttaatt	103440
tactaagcag	aacttcagat	gaggaataaa	atgattggaa	tatttttttt	ctcctctaac	103500
tacttgtaaa	tttgggagaa	tttggagagt	gtagtagagt	cagatcagtg	tatggaaaag	103560
gagcaggagt	gactggacct	tctaagaagt	gtgttatcag	aattagtaaa	tgaaggggtca	103620
aatgtcctac	ttttcccttc	cactgatttt	gacatcaaac	cattatccac	atagccttat	103680
ttcctccctc	ggtcttaatt	ttattaatat	tttactgcac	tttgcagata	aaatttttaa	103740
aaaattttta	aaaattgcca	ataagtgaca	tttattaagt	tcagtgccta	gtgtatatatt	103800
ggattttatt	tattagtcac	aagacctttg	tgcaggtagt	aggcatgatt	atcttttttt	103860
ttttgagatg	gagtccttgc	ctgtcgcccc	ggctggagtg	caatggcgcg	gtctcggtctc	103920
actgcaacct	cggggttcat	gccattctcc	tgcctcagcc	tcccaaatag	ctgggactac	103980
aggcgccctgc	caccacaccc	ggctaatttt	tttgtatttt	tagtagagac	ggggtttcac	104040
catgttcgcc	aggatggtct	cgatctcctg	actttgtgat	ccgcctgcct	cggcctccca	104100
aagtgtctgg	attacaggca	tgagccaccg	cgcccggaact	gattatctta	tttacacatg	104160
agaaaaaccag	ggcttagaaa	ggttaggtaa	cttctcttag	gttgtacagt	aaatgtggac	104220
ctagaagcat	tttgacaaga	gcacctgttt	tttttcttct	tctattagtt	tagaaattat	104280
atactcttaa	ttatcacctg	ggattttgat	tagacagcct	tcagtgtctt	tttcatctta	104340
aatgttcttt	gtgtcttaaa	gggctaagtg	atttcttcag	atcttttagt	tcactcattc	104400
tcagtgaaat	aaaatgaggt	ctaactctgt	actgaatcaa	gttttcagca	tgttattttc	104460
ttctcccttc	ctctctcct	tccttccctc	aaccaggctc	ccgaggagct	gggattacag	104520
gcgcccccca	ccactcctgg	ctaattttta	tatttttagt	gagacggggt	ttcaccatgt	104580
tggtcaggct	gatcttgaac	tcctgacctc	aagtgacca	cctgcctcgg	cctcccaaag	104640
tgctgggatt	acaggcatga	atcaccacac	ctgacggcat	gttattttca	tcgcaaagtt	104700
actgtaagct	gggagaagtg	gcacacactt	gtactcccag	ctactcagga	agcttaaggt	104760
gagaagattg	cttgagccca	ggagttttga	gaccaacctg	ggcaacacag	caagacccca	104820
gctcaaacaa	agaaaaaaaag	ttattgaatt	ttttatttct	atggatcatt	ttttgtagtt	104880
tcttattcct	ttcacccttc	attcccactt	ttgatcccat	cttttattta	tttagtttta	104940
ttaaatgtat	attttgtctga	taattctgct	atctacagtt	ttttgtggac	ctgactcagc	105000
atttctttgt	ttcttcggat	tcagactgtt	ggtaggctgt	gatttttagtg	atttttggcc	105060
gtgaacatgt	ttcttggaact	tttgtctgtg	ggaattctct	gtgtactctg	tataaattaa	105120
gttacttcag	gtgttttgca	ttttcttttg	ccatgcacct	ggggcctggg	tcactaccct	105180
cttggtacca	cttaaaactg	aatttttgtc	ttgggtgtct	gtactgatcc	tgtatgagta	105240
caggtttata	actactgtag	aaatatggtg	tttgattatg	gggtattgtc	ccagatgggt	105300
ctggagattt	aatatgtctc	ctgttaaaact	taattgtgtg	tccctgtaaa	actccaaaat	105360
tctgaattcc	agaatactac	tggcccca	tgtttaagat	aagggcactg	cctgtatttg	105420
tttctgcctc	ccactatttt	ccttagttta	acacaaactc	acctttttta	aaaacatttt	105480
gagagaattc	agtattggga	agagtttcta	acctgtttct	ggaaatggaa	gtccaaagtc	105540
tgtttctgta	attgtttttt	ttttgagatg	gagtcctcct	ctgtcaccga	ggctggagtg	105600
caatgacgta	ctctcagctc	actgcaacct	ccacctcccg	ggttcaagcg	attctcttgc	105660
ctcagccccc	tgagttagctg	ggattacagg	tgccaccac	catgcctggc	tgatttttgt	105720
atttttagaa	gagatggggg	ttcgccatgt	tggccaggct	ggctctgaac	tcctgacttt	105780
gtgatctgcc	cacctcagcc	tcccaaagtg	ctaggattat	gtttctgtaa	ttgtaataca	105840
tttattgttt	ttagaaactg	tctttgcttt	agtggtaatt	ttcaataaaa	atagaaatag	105900
cagtggagtt	attaaaagag	cattagttac	atttttccct	ttttcattat	cttcaaatat	105960
tatatatagt	aagtttgacc	tttttaaaat	gtataactgt	atcagtttta	acacatacat	106020
agattcctgt	aactgtcacc	actataaggg	taaagaacag	ttagttcctt	cacctttgaa	106080
gtcaagccctc	acctctatcc	caacacttgg	caaccgtgta	tctttctccg	tctcaatagc	106140
tttgccctttt	ctcttttttt	ttcttatttt	tttttttgag	acagcgctct	gctctgtcgc	106200
ccgagctgga	gtgcagtgag	gcaatctcgg	ctcactgcaa	cctccgcctc	ctgggttcaa	106260
gcagttctcc	tgcccttagcc	tccctagtag	ctgggattat	aggcacgcac	caccacaccc	106320
ggctgatttt	tttgtatttt	tagtagaaat	ggggtttcac	catgttggcc	aggctgggtc	106380
caactctttg	acctcaagtg	atccacctgc	ctcgccctcc	caaagtgtcg	ggattacagg	106440
cgtgagccac	tgtgcccac	caggactttt	tttttttaaa	tttacattca	acttgtcatt	106500
ttttcttctg	atggattgtg	ccttcagagt	cacacctaa	agccctttgc	ctaagcaaag	106560
gtcatgaaga	ttttctcata	tgtttccctt	taaaagtatt	gtgggtggcc	agggtccatg	106620
gcttatgcct	gtaatctcag	cactttgaga	agctgaggtg	ggcagattac	gaggtcagga	106680

46/122

gatcgagacc	atcctggcta	atgCGGTgaa	accccatctc	tactaaaaat	acaaaaaaa	106740
aaaaaaatta	gccgggCGtg	gtggCGggca	cctgtagtc	cagctacttg	agaggttgag	106800
gcaggagaat	agtgtgaacc	cgggaggtgg	agcttgCagt	gagccgagat	cgcgccactg	106860
cactccagcc	tgggcaacac	agtgagactc	catctcaaaa	aaaaaaaaaa	agtattatgg	106920
ttttacactt	tacgtttaga	tatatatctt	ttttgagtta	atgtcgata	agtatgaggg	106980
ttacgtcaga	ttttttgttt	tttgtttatt	tttacatatg	gatgtctagt	tgttctaata	107040
ccatttgttg	aaaagacaac	ctttactcca	ttgaattgcc	tttgtacttt	tgccatattc	107100
gtctaggcct	gtttttggac	tcctttttct	gtttcatgat	gtgtgtgtct	attcctttgt	107160
taataccaca	tggctttaat	tactgtatag	taagtcttaa	aattgggtaa	tgctggcctc	107220
ataaaacgaa	ttgggaagt	tttattttta	ctcttatttc	cattttctag	aagagattgt	107280
gtagaattgg	tgtcatttct	tctttagata	tttggttgaa	ttgggaagt	atgccatctg	107340
ggcctagggt	tttgtttttt	gtgtgtgaga	cagagtctca	cttctgtcac	ccaggttgga	107400
gtgcagtgg	gagatcctgg	cttactgcaa	cctctgcctc	ccaggttcaa	gttatcctcc	107460
tgccctagcc	tcccaaatag	ctgggattac	aagcgtgtgc	caccatgccc	gactaatttt	107520
tgtattttta	atgcagacag	ggtttcacca	tgtagccaa	gctgggtctg	aacttggtgac	107580
ctcaagtgat	tagccacact	tggcctccca	aagtgttagg	attatagatg	tgagccaccg	107640
tgccctggcag	gggcctagg	ttttcttttt	cagagtattt	taaactatga	attcagatta	107700
tttaaatgat	ataggactat	taaagtatc	tgtttcttct	tgagtgaatt	tttactgtag	107760
tttatggcct	ttgagtaatt	aattgtattg	aattgtcaaa	tttatgagcg	tgtaattatt	107820
tatagcattt	cggttttgta	gtggtatccc	tcttttattc	ctgggtgttg	caattgtgtc	107880
ttgtttttct	ttgtcagatt	gtatagggat	ttatttagtct	tttcaaagaa	ctagcttttg	107940
ttttgatttt	tctgtgtgtt	tgttttcaat	tttattgatt	ttctgctctt	tattatttct	108000
ttctattatt	ttctgtttgc	tttggtttta	ttttactctt	ttttttttct	ccaagttgct	108060
taaagtagaa	acttagattt	ctggtttgag	acctttcttt	tctaagataa	gcattttaata	108120
ctgtaaat	ccttctaacc	actgctttag	ttacaccccc	acaaattctg	gtattttgaa	108180
ctgagcacaa	atgaaatgtt	ctaatttccc	ttgaatctta	ttcttttacc	aatgaattat	108240
ttagaaatat	gttatttagt	ttgcaagcaa	ttggagactt	ttttcctggt	atttttctac	108300
catttatctc	tcatttcatt	atattatggt	cagagaatat	attttgaatg	atttcattta	108360
ttaattttta	aaaataacat	taaaaaattt	tttaaaatgt	gaatatacca	catacagtat	108420
aaagattgta	cattctgttt	ttggacagtt	ttctataaat	gtcaagttga	tttagttggt	108480
taatgatggt	gttcagtttt	tctttattct	tgctgatact	ttgtatgcag	ttatatcact	108540
ttattactca	gaagagtgtt	gaactttcca	actacaattt	ttttttccaa	ttttactttc	108600
agctctatct	ggttttgctt	catgtatttt	gaggctctgt	tgtaggtgt	gtacacattc	108660
aggatgat	cttctgggtg	aattgcctgt	tttatcatta	tgtaattccc	tctttatggt	108720
aattttcctt	gttctaagat	cagaaatata	tggtgtccaa	tttatataga	cactgcagct	108780
ttcatttgat	tagtgttgc	atggcatatc	tttttccatt	tttttacttt	tgatctacct	108840
ttataattct	atttaaagg	ggcttcttgt	aggcagcata	tagttgggta	gtgttattta	108900
tttatttatt	tattttattta	tttatttatt	tattgagaca	gagttttgct	cttggtgccc	108960
aaagctggagt	gcagtggtgc	aatcctggct	taccacaacc	tccacctcct	gggttgCagt	109020
gattctcctg	cctcagcctc	ccaagttagt	gggtattacag	gcacgcgcac	catgcctggc	109080
tgattttttg	tatttttagt	agaaacggat	tttcaccatg	ttagccaggc	tcgtcttgaa	109140
ctcctgacct	caggtgatcc	acctgctttg	gcctcccaaa	gtgctgggat	tacaggcggtg	109200
agccactgca	cccggtgag	tcagtgtatt	tttaatcttt	tctcacaata	caggggtttt	109260
gttggttaaat	ttaatatttt	taataataat	tttagtataa	ttatttacat	taaagttaac	109320
tggtgactg	gggtatttat	aatgtgtaaa	tataattatt	ggtattaata	taattatatt	109380
actcataata	atattaatat	ctttggattt	agattaccag	tttagtatat	gtttttctgt	109440
ttctccctct	ttgatttccc	cttttttgc	tttttttttt	ttttaattct	tatttttttt	109500
tagtatttgt	tgatcattct	tgggtgtttc	ttggagaggg	ggatttgcca	gggtcatagg	109560
acaatagtgt	aggaaggtc	agcagataaa	catgtgaaca	aggtctctgg	ttttcctaga	109620
cagaggaccc	tgcggccttc	tgCagtgttt	gtgtccctgg	gtacttgaga	gtaggagtg	109680
gtgatgactc	ttaacgagca	tgctgccttc	aagcatctgt	ttacaaaagc	acatcttgca	109740
ccacccttaa	tccatttaac	cctgagtgg	aatagcacat	gtttcagaga	gcaggggggt	109800
gggggtaagg	ttatagatta	acagcatccc	aaggcagaag	aatttttctt	agtacagaac	109860
aaaatggagt	ctcccatgtc	tacttctttc	tacacagaca	cagtaacaat	ctgatctctc	109920
tttcttttcc	ccacatttcc	cccttttcta	ttcgacaaaa	ctgccatcgt	catcatggcc	109980
cgttctcaat	gagctgttgg	gtacacctcc	cagacggggt	ggcagctggg	cagaggggct	110040
cctcacttcc	cagatggggc	agccgggcag	agggcggccc	cacctcccag	acggggcagt	110100
ggcgggcg	agggcggccc	cacctccctc	ccggatgggg	cggctggccg	ggcgggggct	110160
gaccccccag	ctccctccc	gacggggcgg	ctggccgggc	gggggctgac	cccccacctc	110220
cctcccagat	ggggcggtg	gccgggcggg	ggctgcccc	cacctccctc	ccggacgggg	110280

47/122

cggtgcccgg	gctgagggggc	tcctcacttc	gcagaccggg	cggtgcccgg	gaggagggggc	110340
tcctcacttc	tcagacggggg	cggccggggca	gagacgctcc	tcacctccca	gatgggggtgg	110400
cggtcggggca	gagacactcc	tcagttccca	gacgggggtcg	cggccggggca	gaggcgctcc	110460
tcctcatccca	gacgggggcgg	cggggcagag	gtgggtcccca	catctcagac	gatgggctgc	110520
cgggcagaga	cactcctcac	ttcctagacg	ggatggcagc	cgggaagagg	tgctcctcac	110580
ttcccagacg	ggggcggccgg	tcagaggggc	tcctcacatc	ccagacgatg	ggcggctagg	110640
cagagacgct	cctcacttcc	cggacggggg	ggcggccggg	cagaggctgc	aatctcggca	110700
ctttgggagg	ccaaggcagg	cggctgggaa	gtggagggtg	tagggagctg	agatcacgcc	110760
actgcactcc	agcctggggc	acattgagca	ttgagtggagc	gagactccgt	ctgcaatcct	110820
ggcacctcgg	gaggccgagg	caggcagatc	actcgcggtc	aggagctgga	gaccagcccc	110880
gccaacacag	cgaacccccg	tctccaccaa	aaaatgcaaa	aaccagtcag	gtgtggcggc	110940
gtgcgcctgc	aatcccaggc	actctgcagg	ctgaggcagg	agaatcaggc	agggaggttg	111000
cagtggagccg	agatggcggc	agtacagtcc	agcctcgggt	ttcacaactt	tggtggcatc	111060
agagggagac	cggggagagg	gagagggaga	cgagggagag	cccccttttt	gctttctttt	111120
ggattattttg	aattttttcct	taaattttatt	tatcttactt	atttatttat	ttttttgagt	111180
gatttctctg	ccacagctcc	caagtagctg	ggactgcagg	catgtgccac	tacaccaggc	111240
taatttttttt	gtattttttag	tagagacagg	gtttcaccat	attggccagg	ctggctcttga	111300
actcttgacc	tcaagtgtac	cacctgcctc	ggcctcccaa	agtgtcggga	ttacaggcgt	111360
gagccaccat	gccctgcctt	tttctagaat	ttatatattg	agttcttgat	tgatctcttt	111420
tatgtaggct	tttttagtggc	ttctctagga	attacaatat	acatactttt	cacagtgtac	111480
tcacatttaa	tattttgtaa	cttcaagtgg	aatgtagaaa	acttaaccac	cataaaaata	111540
gaactagggg	tgagggtaaa	aaagagagag	aaaagaaatg	taataaagat	ttaataacac	111600
cgttttttttt	ttttttttct	tttttttttt	gagacagagt	ctctctttct	gttaccaggc	111660
tgagtgctgag	ttggctgac	ttggctcact	gcaacctccg	cctcctgggt	tcaagtgttt	111720
ctcctgcctc	agcctactga	gtagctggga	ttacagggtc	gcgccaccat	gcccagctaa	111780
tttttgattt	tttagtagag	acggtttcac	tgtgttgccc	aggatgggtc	cgatttcttg	111840
accttgatgt	tcgctctcct	cagcctccca	aagtgtcggg	attacaggcg	tgagccaccg	111900
cgcccggcta	agtctttaa	tatttttttg	acattgcact	ttttctcttt	tccttctagg	111960
atttttagtaa	cccaaatgtt	agttttgtta	ttgtttggca	ggttcctgag	gctttcctta	112020
cttctttaaa	tttttttttc	ctgttggtca	gcttcgaaaa	tttctattca	tctgtcttca	112080
aattcactgg	ttctttcccg	ttatttccat	tctgttattg	agtccttgta	gtgaatttta	112140
aattttgttt	attatgtttt	ttagttctaa	aattttcttt	ttttgtgtat	gtcttatact	112200
ttgtcctga	aactcttatt	tgtttcagga	tgatcttat	ttcttagagc	atggttttag	112260
tagctactta	aaatttgttt	tatcatccca	gcatatgtgt	cctcttgatt	gtctttctc	112320
ttgtgagata	atgggatttt	ctggttcttt	atatgacaat	taattttgga	ttgtatcttg	112380
gacagtttga	cttacgttac	atgattctga	atcttgttta	aatcctgtgg	aaaatattga	112440
agtttttgct	ttaacaagca	gttgacctag	ttagggtcag	tcacaaaatt	ctaagcagca	112500
ttctgtcggc	tctgggtcca	tcatcagttc	agttttgtat	cttatctgct	tatgtgcctt	112560
tctgtgtcca	gtctgggacc	tgcccaatgg	tcaggtccca	aagcctttgt	acacttttag	112620
aagcagggcc	atgcacaccc	agctcacgag	tggccccggg	agtgcacata	caactcgacg	112680
ttttcatggg	ctccttcttt	tctgtgatgt	ccctgacacg	ttctgccttc	taagaacctc	112740
cctttatccc	tttctgttg	tctggctaga	aagtccgggc	tttagattcc	ctatacttca	112800
gcacacttcc	tgtagctatg	tcaacctctg	tggccacgac	ttcttcttct	tgggactgca	112860
gtttctcttg	tcagaaagta	ggattcttgg	agctgctgtc	attgctgctg	tggctgctct	112920
gatgctgcct	gggagtcgaa	ggagagaaaag	gaacaaaaca	aaacaacca	ggggatttcc	112980
tccactctct	ttgatccgtg	agagccccct	ttcctgttcc	tcagaccaga	aatagagggc	113040
ctgtcttgga	acttcttctt	tgtgcactct	gtgtgcagtt	tcagcttttg	agtccaggcc	113100
aggaggtgct	ggacaaactt	gtcaggagta	cggaggtact	gcaagttctg	attacttttc	113160
tcagtccacc	tgcttccaag	tccttggatg	catttggtcca	ttgttttgag	ttgcattcca	113220
tgggagagac	agaagagtgt	gcttatttca	tcttgacata	cttattagga	tttcatatca	113280
aatcaacagg	tgatattctc	tatatataat	tgctgttttc	cctttagcaa	gcacattag	113340
aaaataacac	tttaacaccc	gcctttgggtg	gtttctgtca	taattattaa	tacttgactt	113400
tttttttttt	tttgagacgg	agtctcactc	tgctcttga	ggcattgtcc	ccataaactt	113460
ttggtaaagc	atcaataatt	ttatctttca	tccacacaag	cttcaccata	aatttgatgt	113520
ttattcttcc	attttagcag	aattcatggt	gctccaatag	gggctgtctt	caaactgatg	113580
ttttctcctt	cttagtgctt	cagagttagat	cctgttcaga	tacgttataa	caggttaata	113640
tgagtttatt	ttgggtgtaa	agtactttga	aattcatgca	tagttttttc	atcatatgca	113700
ttttccatag	ctttgaacac	ccccatgtaa	ctctcctctt	ccacaaacca	aacaatgaaa	113760
aagcaccttt	gtgatggaag	tttattttgc	aataggaact	cacagtgatc	taagccctgc	113820
tattcatgaa	tataattcat	tactggagtc	caagttgctt	tttggttttt	gaagttctct	113880

48/122

tcttcccttg	caggtataga	acaagatgca	gtgaatactt	ttaccaata	tatatctcca	113940
gatgctgcta	aaccaatacc	aattacagaa	gcaatgagaa	atgacatcat	aggtaagcag	114000
tgcttgaaac	tatggcaaaa	aaaaaatgac	aaaaaatgca	cagaactgac	aattttcggt	114060
attgactaag	ataatttttt	cttaacatgg	aatttagcag	ttcccttcc	aattttgttt	114120
ctgagtat	tttatatcgg	attatagctc	actttaaaag	tttctcggct	gcattcggtg	114180
cgagggctt	tgcttgggccc	agatgggctg	cagtgtagcg	gggtgctcagg	cctgcccgt	114240
gctgagcagc	cgggcccggcg	ggcggtacg	ctaaccggca	cagaccaccg	gatggactgg	114300
ccggcagccc	cgcaccagtg	cacgaagtgg	gcgggacaga	aacttctggg	gttgggaagtc	114360
cagtgaggct	aaaagccggg	accaaagtct	ctaggcatca	gggctgcagc	ccaagagtct	114420
cagcaccagt	gggcaactgg	atggccagac	aggtgtctca	gtgggtggcct	ctccgtctca	114480
gggcttcatc	ccacttctca	gtgggcctga	cgtccctggg	caccctggat	gtctacctgc	114540
attagccaga	gccatcacat	ggcctgtgac	ttgccttttt	ttgccagtgt	attgtgccac	114600
acacagtgtc	atttctgtgt	catttggcac	agctggagggt	gcaaggagga	gggcagcctc	114660
atgtccagtc	ccagtttcac	gtaactttat	tcttctgaat	aaagacaatt	tgctaacctt	114720
aaaaaaaaaa	aaaaaaaaaa	agtttttctt	atatgttggg	cccaaattct	taggctttta	114780
cctgaataac	aatgacagca	agatcaataa	atagtacaca	tttattaaac	actcactgtg	114840
tcccagacaa	tattccaagc	actttttatg	gatagactca	ttttaacttc	ttaaagaactt	114900
tgtgggataa	atacagttat	tttatagatg	aagaaactga	agcacagaga	agttaagtgc	114960
tttgtccagg	gtaacagctc	agatatggca	gagtcaggat	ttgaaactag	accctcacat	115020
accttaactg	ctgtgctgtg	gcagtggttt	tcatactgta	ggttgggacc	agccttctct	115080
tatgccctca	ccccctgcc	aaaaaaaaaa	aaaaaaaaaa	aaatatatat	atatatatat	115140
atatatatat	atatatatat	aatatatata	tatatataat	atatatatat	ataaaatata	115200
tgtattagta	tatatgcata	tatatatat	attatatatt	agtatatata	ctaatatata	115260
atatacatat	tagtgtgtgt	atatatatat	atactagaat	aaaaaaaaatca	aagtatctca	115320
gagtagtaag	gacaaacatt	tcagaaaaat	gttttctatta	tatatacatg	tatgtatgtg	115380
tatgctgatt	caacaaatat	atttcttata	ggttatagca	aaatagtttg	aaagctttta	115440
ctgtgtttta	tcaggaagac	cttaggtgaa	cgtatatcca	cagataaaaag	agggttattha	115500
ttcattcaat	aaatattaca	ttctcataag	tcctaataatt	atgtattttt	attcttcaaa	115560
aaagttagta	tttgtatttt	atgaaataag	acatgttctt	gcacttttag	cagatctgtc	115620
ccgatgttgg	gcttctttta	tccttagtgt	gggtgctttg	cactcactca	ctgctgggga	115680
cagcaagacc	cctgttagtc	tcagctgtgt	ttcttaaatt	ggcccactgt	accttccagt	115740
tagctattct	ggggtccatg	tcagtgtggc	tccattttct	ttttctttct	cccacacaga	115800
tacctataac	ggctataaca	taggcctggg	ggctgttggg	ggcttatccc	tatctgtctg	115860
tatttaagg	gtactgtttc	actgagtttt	gctgacagat	gttgtcatga	gatttgagggt	115920
tttctgtgtt	gttgctctat	ttttatgtgg	gaatttgcta	ctatcatcat	ccctagacca	115980
gcttttccca	gtaatacaac	agggatgttc	tgactgatta	gagtttgcct	gtttgaagaa	116040
ttgggtggct	agtatttttt	ttttgagggg	gtctgtgacc	agttaatagc	ctgactggcg	116100
tgtgggataa	aaggaagcag	tttcaagtc	aataaaacac	ttaaaatgaa	accacactgc	116160
aactctcttt	cttttactta	agcttaatca	aattaatgat	gatgtaatcc	catgaaggaa	116220
aagtcttctg	aaggatcaag	ttgataacat	tttgtgatca	aagaatttga	gaaaacctct	116280
atcccagtg	ctatcattat	atatttttag	atgttaatta	cctgtgtggc	tttaggcaag	116340
tcatttttcc	tccttgagcc	ccattcttaa	tcctgtccaa	attatttgtc	tcctcttgca	116400
gttggaactat	tttaatatag	ctgtccttca	agtgagtttt	gttcaaagga	gccttcactt	116460
tagctcttac	tgtgtaccca	ctttgcatag	tcttggtttta	aatgtaatcc	ttggattttt	116520
gggtgtgcta	actaattact	gtttttatgt	gaggatttag	agtgatccag	aatctatact	116580
tgcaactacct	ccttcatctt	ccacaaatgt	ttgaagtggg	agaattttta	aaaactttga	116640
aggtacagct	gacagaat	gctgatgggt	tgggaagtgg	tggtatgaga	gggaaaaaaa	116700
ggaataaagc	atgactgcat	tttttgtttg	tttgtttgtt	tgtttttgag	acggagtctc	116760
actctcgcca	ggctggagtg	cagtggcggtg	atcttggctc	acggcaacct	ccgcctcctg	116820
gggtccagcg	attcccctgc	ctcagcctcc	caagttagctg	ggactacagg	cgctcgccac	116880
cagcgcctgg	taattttttt	ttttgtattt	tagtagaaac	ggggtttcac	cgtgttggcc	116940
aggatgggtc	ccatctcctg	acctcatgat	ctactcacct	tggcctccca	aagtgtgag	117000
gttacaggca	tatatataag	catataaagt	gtgttatagc	atacaaacag	gtatatatat	117060
aaacatgcag	tccacacagc	tgataggaat	gaggcagtag	tgaaggagaa	gttgatgtag	117120
gagaggggac	agttgtttaca	ggaaagaagt	ctggaggcag	aagggatgaa	ttccagtgct	117180
cacatagaag	attgcttaga	tgggagcaag	gacaatttat	ctagagtcac	aggaaagaat	117240
gcagtacacg	ggtagagatg	caggtgagtt	gaaagatgtg	agagatgatg	gaaataattt	117300
tctgattgct	tctatattct	caagggaagca	ggaagcaaa	tcctcagcaa	agagaataga	117360
agaggtgtta	aatatttgag	aaaggagatg	tactgtagaa	aaaaaaaaaa	ctcagtttct	117420
ccttctgaac	tctcacaaaa	cagaaccctt	ccatgactct	agttgtgtgg	ggttttttcc	117480

49/122

ctgtcagcta	ccaattctgc	agatgattgt	tcagtgaaca	ccaactgggt	gtcctctaag	117540
tcagttcagt	tctcacactg	tttacctgga	gatagcatca	gatcccacag	attgaggact	117600
ctgtcccaca	agactgcctc	cacttcagat	gccagtctca	agtacaagtt	gtggcctgtg	117660
cttctgactg	accttctata	aattggagtt	cccacagtcc	cctccttggg	ttcaataaat	117720
ttgctagagc	agctctcaga	actcagggaa	atgctttaca	tatatattacc	catttattat	117780
aaaggatatt	acaaaggata	cagattgaac	aggcagatgg	aagagatgca	tgggcaaggt	117840
atgggagagg	ggcacagagc	ttccatgcac	tctccaggtc	atgccaccct	ccaagaacct	117900
ctacagattt	agctattcag	aagccccctt	ccccattctg	tccttttggg	ttttttgtgg	117960
agacttcatt	atataggcat	gattgatcat	tggctattgg	tgatcagctc	aaccttcagc	118020
ccccctcatcc	cgggagggtg	gtgggtaggg	ctgaaagtcc	caaacgtgta	attctgcctt	118080
ggctcttctg	gtgatttagc	ctcatcctaa	agctctttag	aggccacagc	cacaagtcac	118140
ctcatttagcc	ttcaaaagaa	tccagagatt	ccatgaattt	taggcgctgt	atgctaagaa	118200
actggctaaa	ggccagttgc	aatgtctcag	gcctgtaatc	ccagcacttt	gggaggctga	118260
ggcaggagga	tcgttttcagg	ccatgagatc	aaaaccagcc	tggccaacat	agtgaagacc	118320
ccttaaaaaa	aatttaaaaa	ttggccaggc	gtaaatgctc	ttgtctgtag	tctcagctac	118380
tcagaagggt	gaggatcact	gagccctgga	gttgaaagca	gcagtgaagg	atgatcgctg	118440
cactgactcc	ggcttgggtg	acaaagttag	accttgtctc	agaagaaaaa	ggaaaaaaa	118500
aaaactgggc	aaagactaaa	taacatatatt	cacagtatca	cagatttgta	ttgtctagga	118560
aagtgaatgt	aaacagacca	ggacactagt	atgatccctt	ggtttcatga	aggctccact	118620
aaagtcatga	acacaaagtg	agactaggca	tcagtgtata	tggtttttcc	agccatgttt	118680
aacagctagc	taaatagcta	attgtttcgc	tgcagtttat	tttagcagtt	ccttatttta	118740
gcacatttca	tgtttttaaaa	tttctacca	taacatttta	ataaactttt	ttacagataa	118800
cttcacaaat	ccataatttt	ttaagttaca	atcccagaaa	tagaattgct	cattgaaagg	118860
gtatgttcat	ttttaaaagt	atgctagaaa	ctgcccattt	gccttcagaa	aaaggtgttt	118920
gtatccccac	taacactagt	gttagttttc	ttgtgccctt	gctcaagtat	acatattatt	118980
aaaaacaatg	ttgggccagt	ttactagata	aaaggtgtag	tgccctcctt	ttctaattcta	119040
tttgattact	agttagtatg	tatgtctttt	cacgttgggt	atttttatgt	tgttcctttg	119100
tggattgtca	tgtcctttgc	tcatttttct	tttggaaacat	ttcttagtag	tttataagag	119160
ctcttgggtat	tttaaatgata	gtaacctttt	aactgtcatg	catgctgcaa	atcttttttc	119220
tgtttggttg	cctttgtatt	ttgttttttg	agggtttcta	tgtataggaa	ttaaatttta	119280
tgttgttaaa	tccttttgatt	tctgtctttg	catatgtact	tcaaaagact	ttctatttta	119340
agatcaagtg	ttacctgtat	tttcttttag	ttctatttta	aacctcttaa	tttatatgcc	119400
tgtgctgtta	actcccaagt	tgattcacaa	gtgtgtatag	atagtttgaa	tttagtggca	119460
atttaattat	ttacaacttc	ttttgcagca	aggatttgtg	gagaagatgg	acaggtggat	119520
cccaactggt	tcgttttggt	acagtcacata	gtcttttagt	caatggagca	agagtaagtt	119580
agttcatatt	ttcacattgt	gcattcctagg	gaatttgggt	tcattgttag	gaatgggctt	119640
cactcagcta	aaaacaaagt	atttttgaga	atttaaatat	tttggatatt	tacaagatca	119700
tataaagcat	actctatctt	ggttaacagt	ttcttttaaa	tataaattat	gtgaactctt	119760
aaaattttca	ttttcatttt	caatgttaat	atttcctaag	ttaaaataat	ttgttttttag	119820
ttctgaaata	atttggggag	tgattgagtc	tgtagtgtat	atgactatta	gaattgggtt	119880
atttatttta	ataatgcatg	tcttcagatg	gctctcctaa	tttgtagtt	aggctttaag	119940
ctaaatggat	gctatataac	taaatccaca	tagatttgtt	gaaatggctc	cagaggtttt	120000
ttagattttat	tactgctatg	tgcccttaaa	aaaaatctat	tcattctttc	acttaacatt	120060
tatcagaaga	gtgctctgtg	taagacgtgg	ttaggcatag	tgccagtctt	gaaggaagtt	120120
acagcccta	aaaagacata	gggcatgttg	tttgggtact	gtaatatgaa	gtggcatgtg	120180
ttaaatgtca	ggggagaact	acaaagtcac	aaaaagggtg	gagagattac	atacaggtaa	120240
aggaatcagg	aatgacacca	tggggagtaa	ggtagtgttg	acctaggcct	ttaagataca	120300
atagggacag	tatggaaaga	gtatatTTTT	cccacttaaa	ctctttcctt	ggtcgttccc	120360
tcaaattttc	ccttttgtcc	atgtgcaggc	acttttagtg	gtttctgcga	agtcaccatt	120420
tctgtaaata	ccagattgaa	gtgctgacca	gtggaactgt	ttacctgggt	gacattctct	120480
tctgtgagtc	agccctcttt	tatttctctg	aggtaaaagt	tgcatcttct	ttcacactct	120540
attcgagcat	tccagcctct	aaactatcaat	gtcggggccc	tgtctatagg	aaataacaca	120600
gaagagccaa	gtcattttcca	aaaagatgta	tcattgtttc	aagttgtttc	tgatggcaag	120660
agtaatttaa	taatatatta	gagagaacat	gaaaattcaa	tgtattaaat	aactctaatt	120720
ttgagaaacc	taattaaact	actgcatgta	agagagtgc	tgtttttaat	tatttggagc	120780
tattttaaaa	ccacagaatt	tgaaacttgc	ttccagtgca	taaattgcag	accagacttc	120840
agaagagaaa	aaaagtagta	aattttttct	tatgtctatc	atttttactt	tagtcacttg	120900
ataggattgc	ccagtgaaga	agcatttgca	acagacaatg	agtatattaa	tccttttgag	120960
gcatacagtt	tagtataatg	ctctttgtta	ggcttcaaca	agtgaattta	ttttgttgga	121020
aagcaaatga	ctattaagta	gaaagaggat	tcccagctct	acaaagcagt	aatttagaca	121080

50/122

ctcgattctg	cctcttttaca	agaatacagg	tactcagttg	atttgttttc	tcactccctt	121140
tctttgctat	aagttttaaat	caacaatttg	tttaggttaa	tatgtcctca	tggaatggtg	121200
gaaatgatca	gatataaaat	atttggtttg	gttaggttac	tctttatatg	tttgctggca	121260
agguaccaca	aatccagttt	agtataattt	ttactctagt	tcactaaaag	tttgcatcca	121320
gctgtgtagg	tagtgtttgt	ttcttgttaa	cttttttttc	gtctaaaaga	atacttttaa	121380
acttttcaat	ctcaaagtac	tgtaacttgc	tgacaggtgt	taacagaaga	agtagatctt	121440
tttgtttttt	gcttatgacc	tgtattttaa	tatttgagct	tatagattag	agattgtgag	121500
agaaatctgt	ttatagtctt	attttccctt	gtgtattttt	tcttcctagt	acatggaaaa	121560
agaggatgca	gtgaatatct	tacaattctg	gttggcagca	gataacttcc	agtctcagct	121620
tgctgccaaa	aagggccaat	atgatggaca	ggaggcacag	aatgatgcca	tgattttata	121680
tgacaagtga	gttatattga	tagatggatt	cagcagatac	ttattgaaca	tttgatatgt	121740
tttgtggaaa	taaagatgaa	taaaactcag	ctctgttgtc	aaggagctca	caggaggcag	121800
cataaaagct	gcttttatat	ggtgtttgta	aagctttggg	ggttcttaga	acaaaagt	121860
ctgtgtggaa	agggggagg	tatgtgggg	aaacaggatg	gcaatggtgg	tgttcaagga	121920
gtgtttccca	tttgtttgga	tttgtttgga	tcccaaaaga	agaaggggaat	tttgctaccc	121980
agaaatggca	gaaaacaaca	ttctaggcaa	aggcattggc	ccagaagcca	tggaacgta	122040
gggttaagt	gcactttcaa	gaaacttgag	tttagataat	caaaggagtg	gggaataaat	122100
atgaggatgc	tggtactaat	tggaaatgat	tgttaaggga	cttgaatgcc	tattttatggg	122160
tatatataac	ttctgtata	aatctgctca	ggcagcttgt	taattagttt	tttattagtt	122220
ttcactgaaa	atgagaggat	ggaaacatca	tacagttaac	aaaattgaaa	atatctggtc	122280
aggcagatga	tgagcttgtg	gccagctctg	taacgtatgg	tattcttttc	atttaacttt	122340
tcttactctg	taaaaaaagt	aattcgtggt	cgggcacggg	ggctcactcc	tgtaatcaca	122400
acacttttgg	agggcagagg	aggtgaatcg	cttgagccca	ggaatttgag	accagcctgg	122460
gcaacatgtc	aaaacccggc	tttactaaaa	atacaaaaat	tagctgagcg	tgatggcgtg	122520
cgctgttgt	cctagctact	taggggctcg	aggcagaagg	atcacctgag	ccttgggagg	122580
tcgaggtgtc	agtgaagctg	gatccactgt	actccaccct	gggcagggca	gtagagtgag	122640
accctgtctc	caaaaaaaaa	aaaaacaaca	aaggtaat	gttatttgta	tccttaagca	122700
aatgttaag	gggttaactg	gggtatagaa	aaagtccaca	gatgttaggg	tttgaagaca	122760
ctaatagtat	ctaggccagt	gggtcctgaa	cattagtctg	tgggctcttg	ctgggctgtc	122820
tgcataaggaa	tcacctgaga	gcttattaaa	aatagggttt	caggctgggt	gcgggtggctc	122880
acgctatata	tcccagcact	ttgggagggt	gaggcaggcg	gattacttga	ggtcaggcgt	122940
tcaagaccag	cctggccaac	atggtaaaaa	ccctctctca	ctaaaaatac	aagaattagc	123000
caggcatgat	ggcacacacc	tgtaatccca	gctactcagg	aggctgagga	aggagaattg	123060
ctcgagcccg	ggaggtggag	gttgcagtga	gcggagatca	tgccactgca	ctccaggctg	123120
gctgacagag	ggagactctg	tctcagaaaa	aaaaaaaaaa	atagggtttc	agtctgggta	123180
ccggtgggtc	acacctgtaa	tcccagcact	ttgggaggcc	aaggcaggca	gatcacttga	123240
ggtcaggagt	ttgagaactg	cctggccaac	atagtgaac	cttgtctcta	ctagaaaacta	123300
caaaaaaata	actgggcatt	ttgacgggtg	cctataatcc	cagctactag	ggaggctgag	123360
gcaggagaat	tgcttgaacc	cgggaggcag	aggactgcat	ctcaaaaaaa	aaaaaaaaaa	123420
aaaggtttcc	agtcacctg	tctcagaaat	tctgattctg	cagggttgag	gtgtgaccag	123480
gaatctttat	ttttagaaga	cataccagat	aatcttgata	aatagccagt	ttagggatgt	123540
agtctaattt	tcctattttg	caagtaagga	aaataaggcc	cagagaggta	atgattttct	123600
caaagtcaca	gaacaagtta	gtggcagaat	ttggactgga	atgcagttct	taatgttctg	123660
tccagtgttt	attctggtac	agtatgtttg	tagaagggtat	tacgtaagaa	acattgttat	123720
atagatgttg	agataggaag	agtttacatt	tagaaatttg	gtctaaaatg	cctgaacatt	123780
caagtcgtgg	aggagtattg	accaacttac	tcaatacaac	ataggagatt	cacattttgt	123840
tacaaaaatg	ctgatttaaa	aggagagttt	tctttttttt	cttctttttt	attttttgag	123900
atggagtctt	gctctgtcac	ccaggctaga	gtgcagtgac	acgatctcag	ctcactgcaa	123960
cctccacctc	ctgggttcaa	gcgggtctcc	tgccctcagc	tcctgagtag	ctgggattac	124020
aggtgggggc	caccacgcc	agctaatttt	tgtattttta	gtagagacag	ggtttcacca	124080
tgttggccag	gccggctctg	aactcctgac	ctcaagtgat	ccaccaccca	ctgcctccca	124140
aagtgtctgg	attataggcg	tgagccactg	tgcccagcct	gcttgttttt	gtatcatata	124200
tatgcatcat	cataatcatg	cattatcaac	ctttgtattt	ctgtcaggac	atagaaacca	124260
ttagagtgtc	tggaagagag	cctttttttt	tttctcgcat	ttaatgcttt	ttttgggtatt	124320
catttcataa	tcagctttacc	aaaacattac	ctgcattata	ccccatcaag	gtagaaatct	124380
ttgtgttatc	aatattgggt	actccctttc	cacaccgagt	catcagtaag	tcctgttcta	124440
tccaaatagg	tcatatgcat	ctagctcacc	cctcagtgct	gttttggttt	gaatttgtag	124500
atgtttactc	ctgatgcctt	gtagttatga	tgatgtgttc	ttattttatt	ctgtgcatac	124560
aagttctcag	ctcgtctttt	agggaaaaatg	accatgtctt	cctttcctat	aaattccctt	124620
ctatctatca	agtcctcaac	agagaatagg	taccataaaa	tatgtgattg	ttagtttctt	124680

51/122

tgccctcagtt	gtagtctgat	ccttacagct	tttaaaacaac	agtagagttc	accgtcaaga	124740
actaaggatg	gttggcaggc	agatagaaag	gtagcaagtt	gacccaacta	tctctgggga	124800
agtgggaaca	aagaaagggt	acatcagcac	tgatcatcaca	tagctctata	gttctaggcc	124860
tgccaggtca	atcaagtagc	cttgtataag	attctctgga	ggaggtgctg	aaagttgctt	124920
atacttgcta	tggaaatttg	ttttacttcg	gatattcttt	taccataggt	acttctccct	124980
ccaagccaca	catcctcttg	gatttgatga	tggtgtacga	ttagaaattg	aatccaatat	125040
ctgcagggaa	gggtgggccac	tccccaactg	tttcacaact	ccattacgtc	aggcctggac	125100
aaccatggag	aaggtaaccc	agaacttcaa	acgtatcaaa	ctacaagaag	ttttattggt	125160
agaactcata	aaatataagg	tgggaaaacc	aagcagaata	gcacagtgga	aattgaagca	125220
gtccagcaaa	gtgattaaga	gcagaggcct	tgagtctggc	ctggtatgta	cagtcacgtg	125280
ccacataaca	tttttagtcaa	cagtggactg	cggtgtacgat	ggctctgtac	gattataatg	125340
gatcaaagct	ggtagtgcaa	taataacaaa	agttagaaaa	aataaatttt	aataagtaaa	125400
aaagaaaaaa	gaaaaactaa	aaagataaaa	gaataaccaa	gaacaaaaca	aaaaaaatta	125460
taatggagct	gaaaaatctc	tggtgcctca	tatttactgt	actatacttt	taatcattat	125520
tttagagtgc	tccttctact	tactaagaaa	acagtttaact	gtaaaacagc	ttcagacagg	125580
tccttcagga	gggtttccaga	aggaggcatt	ggtatcaaaag	gagatgacgg	ctccatgcgt	125640
gttactgccc	ctgaagacct	tccagtgagg	caagatgtgg	agggtgaaaga	aagtgttatt	125700
gatgatcctg	accctgtgta	ggcttaggct	aatgtgggtg	tttgtcttag	tttttaacaa	125760
acaaatttaa	aaagaaaaaa	aaaattaaaa	atagaaaaaa	gcttataaaa	taaggatata	125820
atgaaaaatat	ttttgtacag	ctgtatatgt	tttgttttta	agctgttatg	acaacagagt	125880
caaaaagcta	aaaaaagtaa	aacagttaaa	aagttacagt	aagctaattt	attattaaag	125940
aaaaaaattt	taaataaatt	tagtgtagcc	taagtgtaca	gtgtaagtct	acagtagtgt	126000
acaataatgt	gctaggcctt	cacattcact	taccactcac	tcgctgactc	acccagagca	126060
acttccagtc	ttgcaagctc	cattcatggt	aagtgcctta	tacagatgta	ccatttttta	126120
tctttttatc	tgtattttta	ctgtgccttt	tctgtatttg	tgtttaataa	cacaaattct	126180
taccattgca	atagtggcct	acgatattca	ttatagtaac	atgtgataca	ggtttgtagc	126240
ccaaaagcaa	taggttgtag	catatagcca	aggggtgtag	tagggccatac	catctagggt	126300
tgtataagta	cactctgtga	tgtagcaca	atggcaagca	gcctaacgga	aattctgttt	126360
attgattgat	tgattgattg	attgattgag	acagagtttc	actccattgt	ccaggctgga	126420
gtgcagttgc	acagtcttgg	cacactgcaa	cttctgcctc	ccaggttcaa	ccaattatcc	126480
tgccctcatcc	tcccaagtag	ctgggattac	aggcaggcac	caccatacct	ggctaatttt	126540
tgtatttttag	tagagacagg	gtttcaccat	tttggccagg	ctgttctcga	actcctgacc	126600
ttaaagtgatc	tgccctgcttt	ggcctccgaa	cattgtcggg	ttacagggc	gagctaccat	126660
gcctgggcag	taactgaaat	tctctaattg	catcttccct	atctgtaaaag	tgacgataat	126720
atgcacgttt	acctcaaagt	tactttgatg	attaaagtaa	ggtaaatgtat	ataaaaataca	126780
tattaacata	gtacctgaca	catggtgaag	atcaaaaaat	gttaactact	tttattacta	126840
ttattattac	gtatttttaa	ataattagag	agcagtagta	aaaattagct	gggcgtagtg	126900
gcattgcacct	atagttccag	ctactcagga	ggctgaaatc	ggaggattgc	atgagcctgg	126960
gaattaaagg	ctgcagtgag	ccgtgttcat	gcccctgcac	tccagccttg	gtgacagagc	127020
aagaccctgt	cttgaacaat	taaagaaggc	attatgccgc	aacgttagct	tagaaatgat	127080
ccacatatat	caccagtaac	tgtcaacagg	attggaaccc	tagttttggg	tattatgatc	127140
acaaggattt	attaatagct	tattaataat	aaagcgttgg	ctaggcacgg	cgactcacat	127200
ctgtaatccc	agcacttttg	gaggccgagg	tgggtggatc	acctgaggtc	aggagtttga	127260
gaccagcctg	accaacatgg	agaaacccca	tctctactaa	aaatacaaaa	ttagccgggc	127320
gtgggtgggtg	atgcctgtaa	tcccagctac	ttaggagggt	gaggcaggaa	aatctcttga	127380
acccggggagg	cagaggttgc	agtgcagctg	gatcgcacca	ttgcactcca	gcctggggca	127440
caagagcaaa	actccgtctc	aaaaaatata	ttataataaa	taaataaaaag	taaagtattg	127500
atgtttgtga	atgattttatt	cttctaattg	actagaggag	atttttccag	gaatttcaga	127560
gccagtggag	ttatgttgct	tgtatgtgtc	atgtgtatcc	agggtgaaaa	acttaattaa	127620
acgtatttat	ataataccat	acataaaaac	tgaattttag	gaatactgaa	gaatgacata	127680
tagaagtcaa	atcattaaat	agctagtagt	aaacagaata	gagtgtcagc	tgttacccaa	127740
tgatgataat	atttttcacga	ttaaaattaa	accttttctg	atttttaaag	aaaagttcag	127800
atctgtatca	tataaagaat	gtaaattttc	agggttaata	aattaaaaatg	cagagagaaa	127860
aatgcaaaaa	tagttcttac	tagatgtgtg	tatgtaagga	acttagacta	attttaagaa	127920
cactgtcaag	accctggtag	ttaggttaga	aaaaagacat	gaatgattca	ttcaacaaaa	127980
actttgagta	tttctgtgct	agatggtagt	gttacagtg	taaacaaaaat	aaatgtgttt	128040
ctgctatcct	ggagcttagt	ctacaaaaaa	ggtacatat	ggccggggcac	ggtgggtcac	128100
gcctgtaatc	ctagcacttt	ggaagatcga	ggcgggtgga	tcacctgagg	tcaggagtct	128160
aagaccagct	tggccaacat	ggcgaaaccc	cgtctctact	aaaaatacaa	aaatttaactg	128220
gggtgtggtyg	cggacacctg	taatcccagc	tactcggggag	gctgaggcag	gagaatcact	128280

52/122

tgaacctggg	agacagaggt	tccagtgagt	cgagatcatg	ccactgcatt	ccagcccggg	128340
ggacaaaagc	gaaaatacgt	ctcaaaaaaa	caaaaacaaa	caacaaaggc	acgtatttaa	128400
tacgaacata	aataatttaca	aattatactg	aataagttct	catgtttatt	atgtgcttgt	128460
ccagttacaa	acttttcctt	cgtagaatta	gaaatataaa	taataaacat	gagaactcat	128520
tcagtataat	taataattat	taaatgtaaa	taaaaacatc	tatgtacaat	taggcattta	128580
tttaagaatt	atttgaaaaa	aaaacaatgt	ggaaacagat	attttgatat	attgctagtg	128640
attgaaattg	ataatgttct	tttgaagagt	aaagtgacca	tatatattaa	agttaaaatt	128700
taactcagca	atcacacgcc	tgggtgagta	tcttaaggaa	atcagtttga	aagtaaaatc	128760
aatatatgca	caaagacttt	aacatttatc	ataaaccaga	aaaatcgagt	ttcaaatatt	128820
atcctatgga	ctattttctg	ctaaaaagta	ttaatatcaa	ctttatgtaa	tactttctgt	128880
acaaatattt	tggggggagaa	aaccacaaca	aattacatgc	attgtaattt	tttttttttt	128940
ttttttttta	gacagtcttg	ctccagcgct	caggctggag	tgcagtgggt	caatctcggc	129000
tcactgcaac	ctccatctcc	cagggttcaag	caattctcct	gcctcaggcc	tcccagtag	129060
ctgggattac	aggcgctcac	caccatgcct	agctaatttt	tatagttttt	agtagagatg	129120
gggtttcttc	atggttgcca	ggctgggtctt	gaactcctgg	tctcaagtga	tccgtctgcc	129180
tcggcctcct	agagtgtctga	gattacaggt	gtaagccact	gcacccagcc	ttatgcattta	129240
taattttaat	ttgtaaactg	tacaaaggga	taatacttgt	agtacaacaa	gaagtaaaaa	129300
catttggtat	aggtagttaa	catttgtaac	cagtagaatt	ataggtaaaa	tttatatttt	129360
taaaacagtt	ttagtggat	ttgatttcaa	ctttaaaata	atgcttttca	tctctatcag	129420
gtctttttgc	ctggcttttt	gtccagcaat	ctttattata	aatatttgaa	tgatctcatc	129480
cattcggttc	gaggagatga	atttctgggc	gggaacgtgt	cgctgactgc	tcctggctct	129540
gttgccctc	ctgatgagtc	tcacccaggg	agttctgaca	gctctgcgtc	tcagggtattg	129600
actgattgct	tctgccatta	gggagaaaaa	catacacatc	ctttccttca	catcccagta	129660
acagatccta	atttttgtaa	attttaagtt	gtggaaaaaa	aagataaaaag	ccaggcagag	129720
tggcctgtgc	ctgtaatccc	agcacttttg	gaggctgcgg	tgggcggatc	acacgaggtc	129780
aggaattcga	gaccagcctg	gccgacatgg	tgaaccccca	tctctactaa	aaatacaaaa	129840
attagccggg	catggtggca	ggcacctgta	atcctagcta	cttgggaggc	tgaggcagga	129900
gaatcgcttg	aacccaggag	gcagaggttg	caatgaacca	aaatcacgcc	actgcactcc	129960
agcctgggtg	acaaagtgag	actgtgtctc	aaaaaaaaaa	aaaaaagaga	gaaataaaaat	130020
tagcctactt	actatcttct	aatcaaagca	tttgtggtaa	cttaaaatat	actgtattgt	130080
aaagtatcat	gctgtttcat	ttaggccatt	attctatttg	aatctgtggc	tgtttctctt	130140
aataaatcaa	gtaatatgga	atataattcat	agcctctgaa	gagctcttta	tgtaagtatt	130200
tatttaggat	actttttgta	aaataagtgga	atgaattctt	aggtctcctt	tttttttctt	130260
ttcttgagac	agggctctct	cgctgcaacc	tggaaattct	gggtctcaat	aatccaccca	130320
ccacagcctc	ctgaatagct	gggactagag	gcatgcacca	ccacgcctgg	ctaatttgaa	130380
attttttttt	ggccaggcat	gatggttcac	gcctgtaatc	ccagcacttt	gggagaccga	130440
ggcaggcaga	tcacgaggtc	gggagatgga	gaccagcctg	gccaacgtgg	tgaaccccg	130500
tctctactaa	aaatacaaaa	attagctggg	tatgggtggc	catgcctgta	atccccagta	130560
cttgggaggc	tgaggcagga	gaatggcttc	aaccaggggag	tcggagggtg	cagtgaagccg	130620
agatcacgcc	actgcactcc	tgcattggtga	cagagtggga	ctccatctca	aaaaaaattt	130680
tttttttaaa	tgatggagtc	ttgctgtggt	gctcaggctg	gtcttgaacc	cctgacctca	130740
aatgccgcct	gcttcagcct	aagtttcttt	tttttttgta	aagagacagg	gtcttgctat	130800
gttggccagg	gtagtctcaa	actcctggct	tcaagcagtc	ctcccacctt	ggcctctcaa	130860
agtgtcggga	ttacaggcgt	gaaccactac	ctataatgtt	gtgtttcact	caaggccttt	130920
tgatttcggt	ttgcattacc	gtgccacatt	gtgcatttcc	ttgacctttt	ttgggttttt	130980
tggagtgcct	tcatatgtta	aaccatacct	gattctcttc	aaaatcacac	aaagtagaat	131040
atcctaagac	aagaaatcta	aggaggcata	aagaagttaa	ctgggttttat	taaactcaca	131100
cagtaaatga	tagagccaga	aatattcccc	ttctagtgtt	cttcaccatc	agcttaattgt	131160
agcataataa	ttttctaat	actgttgaca	aataaataac	cctttgaaat	ttcaatactg	131220
ggccttggat	aaatttttct	aatttgtaag	agagtattat	cgtattgcca	tttacaagc	131280
tctcctgagt	atctttttct	tctgttaagt	ttacctagga	gataaactcg	tgagtattgt	131340
tgccattttg	gttttttgat	atagggttaga	atgtcttggg	tttttttttt	tttttttttt	131400
gtttttgttg	ttgtcattgt	ttgagacagc	atcttgctct	gtcgcccgag	ctggagtgcg	131460
atggcacgat	cgtggctcac	tgaacctcc	acctcccggg	ttcaagcaat	tctcctgcct	131520
cagcttccct	agtagctggg	attacaggca	tgtgcaacca	cacctggcta	atttttgtgt	131580
ttttagtaga	gaaggggttt	caccatgttg	tctcaggctg	tattgaactg	ctgacctcat	131640
gatccacctg	cctcggcctc	ccaaagtgtc	gggattgcag	gcatgagcca	ctgcacctgg	131700
ctgaatgtct	tgtttttgat	taggcactta	agaaaggcct	aggtaactaac	cataaaatat	131760
atttttatac	cttttggtga	tactatatat	atagaaaact	gcacttatca	taaccttaga	131820
caccttgaag	aatgttcaca	agcagaacta	acccatgtga	cccagcatcc	agatcaaaaa	131880

53/122

cagcattatc	agccccctcta	gaagccctct	tgggcccctt	ccattcactg	tccttcttgt	131940
caccagggtg	gctactatcc	tgacttttga	tggcatagat	tagcattacc	tgttcttgtc	132000
atcttataaa	taaaaccata	ctgtgtattc	ttttcttgta	cagctttatt	gtgctaattc	132060
acatttacat	catacaattc	agtgggtttt	atatgggtcac	agagttaggt	aaccattacc	132120
acatcgattt	tagaacattt	ttttcactcc	agatagaaac	cccctttact	taaactccaa	132180
atccccact	ccaccagccc	taggcagcca	ctagtctact	ttttatctct	atagagacaa	132240
tagatttgct	tattctggac	atttcataaa	catggaaccg	tatattatgt	ggtcttttgt	132300
tgccaactgt	ctttcactta	gcatcatgtg	ttcaaaagag	catcatgtta	tccatgtttg	132360
gcatgtatca	gaattttatt	cctcattatg	gccaaatctc	ccatttgcaag	gatttatgac	132420
atcttatttg	aattgtaccc	tcctttctgc	catcttatcaa	taatgctact	gtgaccattt	132480
gtgtacaagt	ttttgtgtgg	atacagggtt	tccttttgg	tttaaatttg	aggtggagtc	132540
ttgtctgtc	gcccaggctg	gagtgcagtg	gcacaatctc	ggctcactgc	aacctctgtc	132600
tcctgggttc	aagcagttct	cctgcctcag	cctcccgagt	atctgggact	ataggcacgc	132660
accaccacgc	ccagctaatt	ttttagtaga	gatgggggtt	caccatgttg	gccagtctgg	132720
tctcnaactc	ttgacctcaa	gtgatccacc	catctcggcc	tcccaaagtg	ctgggattac	132780
aggggtgagc	cactatgccc	ggctgtgggt	ttcatttctt	ttgttgtata	tacataggag	132840
tugatttgct	gagtcgaagag	gtaaactctta	aacttattga	aaaactgcc	gattgttttc	132900
cgaaaggtct	gcaccatttt	gcaatcccac	cagcagtgta	tgagttttac	agcttctcca	132960
catctctttg	gaacttatta	tctgtttggc	tgtttttaaa	aatgatagtc	attccaataa	133020
gttctacttc	agtgtgggtt	ttgcacttct	ctgatgagta	atgatgttga	gcattctttc	133080
atttctttat	tggcctttgt	tctagctttg	gaaaaatgtt	tattcaaatc	ctttggccat	133140
ttctttttt	atctttattt	atcttttttt	ttttgagacc	aagtctcact	ctgtcagcca	133200
ggctgttgtg	caatgggtgtg	gtctcagctc	actgcaacct	ccgcctcctg	tgttcaagtg	133260
atttctctgc	ctcagcctcc	cgagtagctg	ggattacatt	tcaggcacct	gccagcatgc	133320
cgggctgatt	tttgattttt	tactagtgc	agggtttcac	catgttagcc	aggtcgggtc	133380
caaactcctg	acctcaggtg	atctgcctgc	ctaggcttcc	caaagtgtcg	ggattacagg	133440
cgtgagccat	tgggcccagc	ctagattttc	ttttttcttt	ttttttttga	gaaggagtct	133500
tgctcttgtt	gcccaggctg	gagtgcattg	gcacaactct	ggctcactgc	aaacctctgcc	133560
tcctggttct	aagcgatttt	cctgcctcag	cctcccaggt	agctgggatt	acagggtgct	133620
accacacac	ccagctaact	tttgattttt	tttttagagac	agggtttcac	catgttggcc	133680
aggtggtct	caactcctga	cctcaggtga	tccacctgcc	ttggcctccc	gaagtgtctg	133740
gattaccggt	atgagctacc	aggcccagcc	aattttctca	ttatattgct	caggctgggtc	133800
tcaaatcctc	gggttcaagt	gatctcctg	ccttggcctc	ccaaagtgtg	gggagtacag	133860
gcgtgagcca	ccttgctcag	cccctttgcc	cattttttaa	ttagattgcc	tttttatatt	133920
gagtttcaag	agtcctttat	atattctaga	taaatgtccc	ttatcaaatt	atattatttc	133980
caggtatttt	cttcattctg	tgagttgtct	ttcctctacc	ttttaaaaaa	gggtgggtttt	134040
tgtttgttgg	tttttaagat	taaggtctcat	tcactctcct	tctgtgccc	aggtcggagt	134100
gcagtggcac	aatcacagct	cactgccacc	tcaacttctc	gggccgaagt	gatcctctta	134160
cttcagcctc	ctgaatagct	agggccatag	atacacacta	tcacaccag	cttttttttt	134220
ctgttttgtg	agacagatct	tactgtgttg	cccaagttgg	tctcaaactc	taggctcaaa	134280
gtgattctcc	cacctctgcc	tcccagagtg	ctgggattac	aggtgtgagc	cacacgcaac	134340
ctgtcttttc	actattaata	gtgtcttctc	gcttcagcct	cccagtagc	tgggattaca	134400
ggcaccaccc	accatgcctg	gctaattttt	ttgcattttt	agtagagaca	gtgtttcacc	134460
atgttcaccc	ggctgggtctt	gaactcctga	cctcaggtga	ttcacctgcc	atggcctccc	134520
aaagtgtctg	gattacaggc	gtgagccact	gcacccggcc	aaaatattgc	cttcttaaca	134580
gtattgtctt	ctaattttgtg	aacatggatg	tatcttcacg	tatttatgtg	ttctttcatt	134640
tcagcagaat	tttgtagttt	tcagagtaga	agcctttcac	ctccttgggt	cattttattcc	134700
tatgttttaa	gttcttttcg	attccattat	aaatagaatt	gttttcttaa	tttcattttc	134760
agattgtttg	atgagagagc	atagaaatac	aagtgtattt	tacatgttga	tcttgcaact	134820
tcaactttga	taaactctgat	tgtagctct	aatagttttc	ttgtggattc	tttaggattt	134880
tcaatatata	agatcatgtc	atcttatggat	agagatagtt	ttttttctgg	ctagaactta	134940
cagagcaatg	atgagtagaa	gtggcagaag	caaaaatctt	tgtcttgttt	cctatctgac	135000
agggaaagct	ttcagtttca	tcatttaata	tgatgttagg	tgtgggtttt	caataaatgc	135060
ctttttttcag	attcaggaat	ttccctatca	ttcctgattt	tttaaggctt	tttttttttt	135120
ttaaatcatg	aaagggtgtt	gaatattgtc	atgttctttc	tgtatcagta	taaatgatcc	135180
tatggatttt	gggtttttatt	ctgttgatgt	gaaatattaa	tgatttttca	gatgttaaac	135240
caaccttgca	tacctgagat	gaatctcact	tggtcatggt	gtataatctt	ttcaatatgc	135300
tgctggattc	cattttactgg	tattttgttg	aagattttgt	atctgaacgc	ttaagataac	135360
atttacactc	tatcagaaat	gaattgacca	taaatgtgag	agtgtatttg	tgggttcttg	135420
attctcttcc	attccaaaga	tagacataca	tccgtctgta	tgtctgtctt	tatgccagta	135480

54/122

ccataactctc	ttgattacta	ttgcttttga	ataagttttg	aaatcagaaa	gtataaatga	135540
gatttttggt	tctgagtaac	agtcctcata	gaattagttg	ggaaatattc	cctcttttatt	135600
ctggccctc	tttctttttt	gtttaactgt	gtatcttgg	gattgttcct	tctcaacaca	135660
tgagagccgc	tttccctacc	ctcccacccc	tgctatagag	aggctctataa	gtgtctgttc	135720
aattatttta	tttacttaac	ctattactta	gtcggggaca	ttaagcttgt	ttatgtcttt	135780
tatttttaaac	aatgctgcag	tgaataatct	tgatataaag	tcattttcca	tcaatataag	135840
tctctctgta	actgaatttt	tagaagtgg	atttctaggt	caacctatgg	ctctgtattt	135900
cacaaaaata	ccaattctgg	tttttcttgt	ggaggtgggg	agtaggaggt	agaatgctgg	135960
aggagaactt	gctgtactca	gctggctagt	catttttagaa	aggtttcctt	agcttctttt	136020
tgatcatatg	cctcaccaag	aatcaaaaac	attcctatgt	accctgtaaa	catggggctt	136080
tactacccaa	gatacatatt	tctggatgta	tgacagcttt	tcataattgaa	gaaataatgc	136140
tgtgagtaca	gcacatttgt	tggaacttag	gtcgttaaga	atgtcttata	aattcataca	136200
ttatacatat	tatttttattt	tatttttttag	tttttgatac	agagtcttcc	tctgtcgccc	136260
aggccagcgt	gcagtggtag	aatcttggct	cactgcgacc	tccatctcct	gggctcaagt	136320
gattctcatg	tctcagcctc	cagagttagct	atgggttacag	gcatgcacca	ccatgcccgg	136380
ctaatttttt	tatttttttagt	agaaactggg	tttcaccata	ttgacctatgc	tggcctcgaa	136440
ctcttggcct	caagtgtatcg	gcctgcctca	gcctcccaaa	gtgctgggat	ccttgtattg	136500
ggttaaaagt	gaatatgtag	ggctgcatgg	tggctcatatc	ctgtaatccc	agcactttct	136560
gaagttaggg	tgggagggag	cctggagggc	aggaggggtga	ggctgcagtg	agttgtgatc	136620
gcqccattgc	acttcaacct	aggaattata	ggcttcagtc	actgtgcccg	gcatgtacat	136680
tttaaatattg	tgctttcctc	tttttagctat	agtatgaggt	tacattttcag	agtcatttgt	136740
gttaagcatc	ttaatagtga	tgaggttgag	tgaaagttag	ttctattttca	aacactgaag	136800
aaaattttgt	acaaatctgt	cacattccaa	gccccaggact	gattgtttca	tatactttcta	136860
attttacaat	ttctattgta	gtccagtgtg	aaaaaagcca	gtattaaaat	actgaaaaat	136920
tttgatgaag	cgataattgt	ggatgcggca	agtctggatc	cagaatcttt	atatcaacgg	136980
acatatgccg	ggtaagctta	gctcatgcct	agaattttta	caagtgtaaa	taacttttga	137040
tcttttaaat	tttttaatta	aatttttacat	ttttttctaa	tctattatta	tatgcccgga	137100
actttcactt	agagtgtgca	gtataatgtg	gtgggttaagt	ataaaggctc	tggagtgtat	137160
tcttgggttt	taatcttggc	tctgccattt	attggcagcc	gctaacctct	tgggtatctca	137220
gttctctcat	ctgtaaaatg	agaataataa	agtgaataa	tgccaacatc	atttactctg	137280
ggctgcataa	ctgatacttg	gaaaaagtat	tccttttagt	ttaagaatta	agttgggtat	137340
tcatttttag	ttgtataata	aagatagtga	ttcataggat	atgccactta	ctgaaattta	137400
ccacagatcc	aatcataaaa	tcactttctc	ttccctaagg	atagcttgat	taacatgtaa	137460
agggtgtgta	aggcttgatt	acactaccct	gatccgtacc	ccagttccca	gcagcaccat	137520
gaaaaaggga	tttcaacata	tttaattact	ttcagtagaa	agtaacagtg	gtaggccagg	137580
cgagtggtg	cacacctgta	atcccagcac	tttgggaggg	cgaggtgggc	ggatcacgag	137640
gtcaggagat	tgagaccatc	ctggctaaaca	cgatgaaacc	ccgtctctac	taaaaaata	137700
aaaaatttag	cgggcatggg	ggcaggcacc	tgtagtccca	gctacttggg	aggctgagac	137760
aggagaatgg	cgtgagcccc	ggaggcggag	cttgtagtga	gcttagattg	tgccactgca	137820
ctccagcctg	cgagtgagg	cgagactctt	gtctcaaaaa	aaaagaaagt	aacagtggtg	137880
ttgggagact	gaggagccta	gaaagtactt	gaaggaaata	aaagggttgt	ttgaccacat	137940
tgattttgga	aagccagctt	tttcagctgt	gtcagctttg	tgtagtgatt	tttagttctt	138000
cttttagaaa	ataacggaca	aggccgggca	cggtggctca	cgctgtaat	cccaccactt	138060
tgggaggccg	agacgggccc	attacctgat	ctcaggaggt	cgagaccagc	ctgggcaaca	138120
tggtgaaacc	ccgtctctac	taaaaataca	aaagttagcc	gggcgtgggtg	gcgtgtgcct	138180
gtagtcacag	ctactccgga	ggctgaggca	ggagaattgc	ttgaaccggg	gaggcggagg	138240
ttgcagttag	ccaagatcac	accattgcac	tgacagctgc	gcgacagagt	aagactctgt	138300
ctcaaaaaat	aataataaaa	taaaaaagaa	tggaacagtaa	acctaataatga	gttcattccc	138360
aaagatgatg	ttattcttaa	gggatgggtc	atttatttaa	gaccttacat	aaagtctatc	138420
aattgcgtga	tttttcaatt	ctgtaattgt	gtgtatgtat	aatgtaaaata	tatatgtttt	138480
tgttttgttt	tgggtttttt	agacggagtc	tcgctctgtt	gctcaggctg	gaatgcagtg	138540
gtgcaatctc	agctctctgc	aacctctgtc	tcccagggtt	aagcgtttct	tctgctcat	138600
cctcccaagt	agctgggact	acaggcacgt	gccaccacgc	ccggctaatt	ttttgtattt	138660
ttagtagaga	tgggggtttca	ccgtgttagc	caggatgggtc	tcaatctcct	gacctcgtga	138720
tccaccggcc	ttggcttccc	aaagtgttgc	tattacaggc	atgagccacc	acaccagca	138780
tgtatttttt	aaatgtataa	aatgaagcag	aaaagagaaa	tgataatttt	tcttcattct	138840
gaaagattat	cttcaccagg	cgcagtggct	cacacttgta	atcccagcac	tttgggaggg	138900
ctcggcaggc	ggctcacttg	agttcgaaac	cagcctggcc	gacatgggtga	aactccgtct	138960
ctactaaaaa	taaataaata	aagatgggtt	taatatatgt	tttagtttta	tgatttttagc	139020
atctttctga	aatttttctc	aaggcaagta	aatttgtatc	agttgggtata	ttgggtaccca	139080

55/122

tctatgaaat	aacttattag	gaagatatct	ctaaaataag	atcactttgc	ctaaaataaa	139140
ctgatataatt	gatgttcaca	gaatttttct	tttaaccgac	ttgataaatg	cattattcctt	139200
gacgtcaagt	gatccacctt	cctcagcctc	ccaaagtgtc	gggattacac	acatgagcca	139260
ccgcacctgg	cattattcctt	ataaaagggt	aaattttctag	ttaagttaa	tgtcctcctt	139320
gttcatgtac	cattgtctat	tttcttccct	tcctactcac	agtaatcatt	ccttatgggtat	139380
gcacttttgt	ttgcttattt	ttatgtaatt	gatattacgc	tccattctgt	acgttgtact	139440
ttcattcaca	gtgagttttg	gacattccta	tgttcatcta	tacagactta	cctcatttta	139500
actacactgt	agtattccgt	atgtaatat	tactataact	catcactgta	gcagagcatc	139560
tcatagtgt	tgtattactg	ttttgccatt	ttgggtatcaa	tgagtattta	agtcattttgc	139620
agtttttccc	tcttataccc	agtattacag	aggatctcct	tttatatgct	tctttgtacc	139680
aagaggcaga	ttaaaaaatt	tttttttgaa	aaaatttttg	aaaaaaaatg	aaatgaagtc	139740
tcaactatgtt	gcccaggctg	gtctcaaaact	cctaggctca	agcaatcctt	ccatcttggc	139800
ctcccaaagt	gctgggggta	caggcatgag	ccaccatgcc	tggcctacat	tttaaatctt	139860
gatagctcct	acaatttact	ttgtaaaagta	tctgcatcat	tttatgttct	caccagtctt	139920
taataagaat	acttcatact	tttggtcgga	cacagtggct	cacgcctgta	atcccagcac	139980
tttgggaggg	cgaggcgggc	agatcaagag	atcgagacca	ccctggccaa	tatggtgaaa	140040
ccctgtctct	actaaaaata	caaaaattag	ctgggcgtgg	tggcgcaccc	gtagtcccag	140100
ctactcgaga	ggctgagaca	ggagaatcac	ttgaacccgg	gaggtggagg	ttgcagtga	140160
cctagatcac	accactgcac	tccagcctag	caacagagtg	agactctgtc	tcaaaaaaaaa	140220
aaaagaatac	ttcagactta	atttttttct	cagctttaag	tgtttgctaa	tgagattgag	140280
tttcttttgg	tatgtctctt	gattgttcag	gttttttctt	ttatgaattg	actgttcac	140340
tctttttcac	attatttctg	ttgggtgatt	ttattagtga	cttgtaaaaa	ttctgtatat	140400
tttttcagca	tgacacttca	ttattcaaaa	aaaaaaaaaag	attctctatg	tttctcgata	140460
ctaactcattg	gttggttaata	ccttaaaaaat	aagaccctta	ctgtattttt	tgcttttttt	140520
tttttttttt	tttttttttt	tttgagatag	agtcttgctc	tgttgcccag	gctggagtgc	140580
aatgggtatga	tctcggctct	cagctcactg	caactgcaac	ctctacctcc	ctgtttcaag	140640
caattctcct	gccttagcct	cccaagtagc	tgggattaca	ggcatccacc	accacaccca	140700
gctaattttt	gtatttttag	tagagacagg	gtttcaccat	gttggccagg	ctgggtctcaa	140760
actactggcc	tcaagtgatc	cgctgcctc	ggcatcccaa	agtactggga	ttacaggcat	140820
gagccacagt	gcctagccac	tttttgcttt	tttaactttgt	tttatagtac	tatagtttta	140880
gtataaacag	atgtatgtat	acacacaact	atggctttat	aatatgtttc	agtcattgtt	140940
agagcaaggc	ctaccttttg	ggtgcttctt	ttacaaaatt	gtcttggcta	ttcttgtgcc	141000
ttttttctta	tttgttaatt	ttagaattgt	ttgactcacc	ttgactcacc	atggtttgta	141060
aactgaggat	tttgaaatgga	attgcactca	attaaagatt	atcttgcttt	ctgtgcagca	141120
atgttttatt	tcaaataatc	cctactttta	attacttagg	atagctataa	attgtgtttc	141180
tggctttcta	gattttagatg	aaacgcttta	aattgattgt	tttctcctaa	atttaaaact	141240
tattgtttaga	agttaaagtc	ttctgttcat	tcttatttag	gaagatgaca	tttgggaagag	141300
gcagtgactt	ggggcaattc	atccgagaat	ccctgactga	accgatgta	aggaaatcaa	141360
aagggttgtg	gtgtttttat	acttcataat	aagcctttac	tcacattagt	gattgactgt	141420
aagtcaaaga	ccacttaagg	tttaaaactgt	ttattttgta	aagtaaccac	tgtatctttc	141480
accttgtgtt	tatagtca	agtaagtaca	agggcttctc	gtagtacat	ctttatgcaa	141540
tctcctctga	atcaaaaagt	agtgaacttg	ctttgccact	ccagaaggca	catgaatag	141600
aaaaagcatt	gtctattttc	ttatttaatg	gcaaaaatacc	cgacctaaat	tggacttaat	141660
gtttgagacc	gtttatttta	ttaaattata	ttttttctct	tttctttttt	ttttttgaga	141720
cagttcttgc	tctgtcaccc	agaccggagt	gcagtggtct	gaccgcacct	cactgcaacc	141780
tctgcttctc	aggttcaagc	gattttcctg	cctcatcctc	ctgagtagct	gggactacaa	141840
gtgcgcacca	ccacacctgg	ctaatttttg	tatttttagc	agagatgagg	tttcaccacg	141900
ttggctaggc	tgggtctcata	ctcctgacct	caagcaatcc	atccgccttg	gcttcccaaa	141960
gtgctgggat	tacaagtgtg	agccaccatg	cctggcctta	ttaaattatt	tttattaaat	142020
ttcctcaaga	ttgatgaaag	taatgaaata	taaaagtaat	gaaatatatg	tggaaaatag	142080
actggattaa	gaaaatgtgg	cacatataca	ccatggatac	tatgcagcca	taaaaaagga	142140
tgagttcatg	tccctttag	ggacatggat	gaagctggaa	accatcattc	tgagcaaaact	142200
gtctcaagga	tagaaaacca	aacaccgcat	gctctcactc	ataggtggga	attgaacaat	142260
gagaacactt	ggacacaggg	tggggaacat	cacacgctgg	ggcctgtcgt	ggggtggggg	142320
gctgggggag	gaatagcatt	aggagatata	cctaataata	atgacgagtt	aatgggtgca	142380
gcacaccaac	atgggtacatg	tatacatatg	taacaaaagct	gcacgttggtg	cacatgtacc	142440
ctagaactta	aagtataata	aattttaaaaa	aaataaatat	atgtggaaaa	tattaatagg	142500
tcaaaattca	aattgttcat	ttaatcagaa	gagtagttta	gtcaaatcca	agggttagac	142560
aacagaaatc	ttttttgtca	agtgcatctc	ttgtgactga	tttcattttc	ttcctgggtt	142620
acacaggaag	atttcagaaa	caaagtgtgga	tccgtgacag	atggtatcta	gaagttttta	142680

56/122

gtttggttga	attgacagta	ttttattgag	taaaagatac	taatttttgt	aagaagaaaa	142740
attcaatttt	gataagtatg	tttaagatta	agagctattg	gccaggcgct	gtggctcatg	142800
cctgtaatcc	tagcactttg	ggaagctgga	gcagggtggg	cacgagggtca	agagattgag	142860
accatcctgg	ccaacatggt	gaaaccctgt	ctctactaaa	ttagccaggc	gtgggtggcac	142920
atgcctgtgc	accgcctccc	gggtttaagc	gacccactctg	cctcagggtc	ctgagtagct	142980
gggattacag	gcgccatggc	taatttttgc	attttttagta	gagacagggt	ttcactacat	143040
tggccaggct	ggtctgggtct	caaactcctg	acctcagggtg	atctgcccgc	cttagcctcc	143100
caaagtgtcg	ggattacagg	catgattcac	catgtctggc	catttatctt	attttctttt	143160
ttttttttt	ttttgtttga	gacggagtct	tgtgtgtgtg	cccagagctg	gagtgcaatg	143220
gtgcgatctc	agctcactgc	aacctctgcc	tcctgggttc	aagcaattct	cctgcctcag	143280
tcttccaagt	agctgggatt	acaggcgctg	gccaccacat	ctagctaatt	tttgtatttt	143340
tagtagagac	agggtttcac	catgttggtc	aggctgggtc	cggaaactcct	gacctcgtaa	143400
tctgcccacc	tcggcctccc	aaagtgtctg	gattacaagt	gtgagccact	gtgcccagcc	143460
atcttatttt	cttctttttt	ttttgtcggg	tgggaggggg	acagagtcta	gctctgtcgc	143520
caggcttggc	tctactgcaac	ctctgcccc	caggttctag	caattattct	gcctcagcct	143580
cccaagtgc	tgggattata	ggcacctgcc	accacgcctg	gctaattttt	tgttattttt	143640
agtagagatg	gggttttgct	atgttgacca	tgtgtggctc	aagtgatccg	cccaccttgg	143700
cctcccaagg	tactgggctt	acaggcgctg	gcttgtattg	ggtaaaagaa	caatattggg	143760
ggctgcattg	tgggtccatc	ctgtaattct	agcactttgt	gagactgaga	tggaggaggt	143820
gttgaggccc	aggagggtga	ggctgcggct	gcagtgaatt	gtgatcacgc	cattgcactt	143880
ccacctaggt	aatggagcaa	gacctgtctc	ctaaaaaaca	aaacacaatt	tttttaaggga	143940
atactgggaa	gagggtcagt	gtgggttttag	aacagaggaa	gtgccagatg	acctttgtga	144000
ggcattggcc	aggaagaact	ctacagtgtc	tttaggtagc	ttctgtccat	aaggataatg	144060
gggtctcctc	cccagtatta	atagaaaatc	ctagagctgt	ttttttttgt	ttgtttgttt	144120
tgtttttttt	tcctgagatg	gagtcctctc	ctgtcggcca	ggctggagtg	ctgtggcgcg	144180
atcttggttc	actgcaagct	ctgcctccca	gggtccacacc	attctcctgc	ctcagcctcc	144240
caagtagctg	ggactacagg	tgtccaccac	cacgcccagc	taattttttg	ttattttttg	144300
tagagatggg	gtttccacct	gtcagccagg	atggctctga	tctcctgacc	tcgtgatccg	144360
ctcgccctcg	ccttgcaaaag	tgtctggagt	acaggcggtg	gccaccgtgc	ctggcctggt	144420
ttttttgttg	ttgttattta	tttattttat	tattttattt	ttgagacaga	ctctcgctct	144480
gtcggccggg	ctggagtgta	gtggcacgat	gtcggctcac	tgcaagctct	gcctgccagg	144540
ttcaagccat	tctcctgect	cagcctcctg	agtagcaggg	accacaggcg	ctcgccacca	144600
cgcccggtta	attttttgta	tttttagaag	agacggggtt	tcaccgcatt	agccaggatg	144660
gtctcgatct	cctgatgtcg	tgatccgccc	acctcggcct	cccaaagtgc	tgggattaca	144720
gggtgtgagcc	accgtgcctg	gcctgatatt	tttttttttt	taatctgggt	tcatacctct	144780
gacagctcat	gaagaagtgc	tcctgcttca	tatgtatatg	tgttagcata	gtgttaacat	144840
agcatagggt	ttcggtgttt	gcagtttctg	ttgtttttat	atgaattaaag	gtgtattatg	144900
agcagttgaa	gatatatagg	aaattttttc	ccaaaccact	atctctgctc	gttctattca	144960
ttcagtcctgt	ttatgttatt	ccttcattca	ttcattttat	agaacagtgg	agtgcctact	145020
gtatgcattct	attgttcttg	gtcctgggga	agaaaaacaaa	gttcctgctt	tcattggaact	145080
tacattatat	tggcggagac	agtaacagac	aaacaaatgt	agcctgtgta	catgtgttac	145140
atgaaaagca	gggtaggggg	ctgggagaga	gtagtgggga	gtgctatttt	cgagggtgggt	145200
gtcaggaaaag	gcctcactga	ggaggtggca	ttttgagtag	acctgagcgc	agcggggggc	145260
taagcccagg	cagcatgtgg	aggaagagtg	ttcttggtga	aaggaacaag	gatagaggcc	145320
cgaagctaga	gagctcagca	tgatcaagga	acagcaagcc	ccgtgtgggt	ggaatggagt	145380
gagcaaaagga	atgagcagta	gaaggtgagt	gagttgggag	gtcaccagag	accatggcaa	145440
ggacttgaaa	gtgtcagggg	cacattggaa	gttggaagcag	ggaaatgatg	ggatttatgt	145500
tttgtttttg	ttttatgttt	agtgttttta	agggattgct	ctatcagcta	tttggaaaat	145560
ttagtgtagg	gcttcaagaa	gagaagcaga	gaaacaacat	tcttgccata	gtcatagtct	145620
aagttaaggga	tgtatgggtg	gtggattagg	ctggtagtgg	aagaccagtc	cagttcgggt	145680
tgtatttgaa	ggtagaggca	aaaagattat	atttctacca	gcaagcccat	ctatgaagtt	145740
acttgattaa	ttaatttaat	tgagacatgc	ccacataaac	taataaatag	gaattttctg	145800
agtttggtta	aacaccccctg	tatatcctgg	ttcttctttt	agttgtccag	atgtctcttt	145860
aagtcaagta	ttttttggtg	gtgtaggagc	ctagagattg	aatttattca	cccaaaaggc	145920
atttgagtga	ttactatgtg	ccaggcacta	tgtctgaatgc	caaggatgta	aataagaggg	145980
cgtagtctca	gtctgtttta	ctccagcttg	ttctcttttt	aatgaccctg	acttggttaag	146040
catatcagtt	atcctacaga	atgtttaatc	ttctgtactt	tcctgggtgt	gttatttagc	146100
ttatttctct	ttccttgaca	tttcttgtaa	actggaagtt	acacctatag	tcttgatgat	146160
tcgtgttaca	catttttagat	tagaacacat	catgtgttgt	atatggtgtt	tttgaaagcc	146220
tctctgtata	ttgggtctgta	cattaaaatg	ttgcctgaat	ggatacacat	aaaatttaac	146280

57/122

agtgattaca	ttagagatga	gaagaaagag	gtgcctttta	cttttcaata	taccttttcc	146340
tctgcttttt	gaacttttct	gccctatgca	tacgtttattg	cttaatcatc	cacctcatct	146400
cttccctgt	ggctttctgt	tgcatttgga	atgaaatcta	gcctctttgc	tgttacctgt	146460
ggatgtccct	tgctggcctc	tatcacctta	ctttgaacca	ctcctttcat	ggactgagct	146520
ctcattggac	tatcttttat	tcttttgctg	aagtttcttc	actttgagtg	cctctgcagt	146580
tgctatttca	tggtgtggc	aagccctgcc	atggctttca	tgcaaggatg	gttctctctt	146640
ctcatctcaa	tattatctct	tcagagaggg	accttcccaa	ctccgatgat	ctaaaaatcct	146700
ttgtatatac	cactcactac	cacttctttc	ttttcttttc	cttttatctt	tttttttttt	146760
tttttttttt	gagataggg	cttgctctgt	tgcccagggt	ggaatcacga	ctcactgcag	146820
cctcatcttc	ttgggtcaa	atgatcctct	cacctcagcc	tctcgagtag	ctggaactgc	146880
aggcacacac	caccatactt	ggcttattat	tttacttttt	gtagagacag	ggtttcacca	146940
aggctgggtc	caagctcctg	ccgcaagcaa	tccacatctc	tcagcctccc	aaagtattgg	147000
gattatagga	gtgagccact	actcctggcc	tattttctta	ttcactgtct	aaaattatct	147060
tgttcattta	tttacatact	tgtttatagc	ttattttctc	gctggacatg	gtgcctcaca	147120
cctgtaatct	caatactttg	ggaggctggg	ttggagaatt	ggttgagccc	aggacttcaa	147180
gaccagcctg	ggcaacaaag	tgagaccctg	tctataaaaa	attgtttaaa	aattagctgg	147240
gcattggtggc	acatgcctgt	gggtcccagct	acttgggagg	cagaggtggg	agaatcgctt	147300
gygccagga	ggttgaggcg	acggtgagcc	atgattgtgc	cactgcactc	tagcctagtg	147360
acugagttag	accatgtgtc	taaaaagtaa	ataaaaaatag	tttctctttc	atgactagaa	147420
tattacctct	atgtgggcag	ggagtttgtc	tatactattt	ggcactatat	ttcctgattc	147480
tgaattatg	cctagcacat	ggtaagtact	ccttaaataat	ttattgactg	aattatttaa	147540
tacttaagaa	tttcatittg	gattatctga	gtggtaagat	tacggattat	atttatgtaa	147600
gaaaaaatca	ttttttaaac	ttggttgccc	tttgccacac	tgacatagac	actaagtttt	147660
cttagccaga	tacttccga	ggatactcac	agaggccatt	ctcttctcaa	tcccaaata	147720
attgatattt	cttagcactt	tcaagctaatt	gcaattctta	gatgatgtat	ctgtgtatat	147780
catatcctca	ttctacaaat	gtagaaattg	aagtctgggc	acagtggctc	tcacctgtaa	147840
tctcagcagt	ttggggaggcc	aaggcgagcg	gatcactgag	gacaagagtt	aagaccagcc	147900
tgcccaacct	ggtaaaagct	tgccctctatt	aaaaatacaa	caattagggc	cgggcgtggt	147960
ggctcacgcc	tataatccca	gcacgttggg	aggccaaggc	aggcagatca	cgaggtcagg	148020
agttcgagac	catcctggct	aacacagtga	aaccccatct	ctactaaaaa	tacaaaaaat	148080
tagccaggca	tggtggcacg	cgcttgtagt	cccagctatc	gggaggctga	ggcaggtgaa	148140
tcctttgaac	ccggggaggcg	gaggttgcaa	tgagctgaga	ttgcaccgct	gaactccagc	148200
ctggtcaaca	gagggagact	ctgtctcaaa	aaaaaacaatt	agccaggcgt	agccaggcgt	148260
gggtggcgggt	acgagtacct	gtaatcccag	ctactagggg	ggctgaggga	ggagaaatcac	148320
ttaaacccag	gaggtggagt	ttgcagcggg	ctgataatgc	accactacat	tccagcctgg	148380
gcaacagagt	gagactctgt	cttaaaaaaa	aaaaaaagaa	agaaagaaat	tgaggaaatgt	148440
ggagattgtg	gtctgtgatt	tgtaggaatt	cacacagcag	gttagtagca	actacagggc	148500
tttgggttcag	aataccacct	tgacaattggt	ttgtttacag	ttcggtctcc	cttctctctg	148560
ctttctctcc	ttccttattg	agggcagctg	gaaagaattt	tcattcattta	ctagcctata	148620
gctttaattt	gagttttgaa	accttgataa	tagagcacag	aggaaaagac	tgagttttct	148680
tttttttgaga	cagtcttgtc	ctatggccca	ggctggagtg	cagtgcacac	atctcagctg	148740
gttgcaacct	ctgcctccca	ggttcaagca	attctgcctc	agcctctcga	gtagctgaga	148800
ttacaggcac	gtgtcaccac	gcccagctaa	ttttctgttt	ttgtttcggt	ttgttttttt	148860
ctgagatgga	gtcttgctct	gtcaccaggg	ctggagtga	gtggtgcgat	gttggtcac	148920
tcaaacctct	gtctcctggg	ttcaagcaat	tcttctgcct	cagcctcccc	agtagctggg	148980
actacaggta	cgtgccacca	tccctagttc	atttttgtat	gtttagtaga	gatggggttt	149040
cactatgttg	accaggctgg	tctcgaactc	ctgatctcag	gtgatctact	cgtctcagtt	149100
tcccaaagtg	ctgggattat	tggtcacacgc	ctatttttgt	atttttagta	gagacggggt	149160
ttcaccatgt	tggttagact	gggtctcaaac	ttctgacctc	aagtgatattg	cccgcctcag	149220
cctcccaaag	tgctgggatt	acaggcgtga	gccaccgtgc	ccagccaaga	ttgagttttg	149280
aaaagagcct	tctgagatta	tgagaagggc	aagcaagata	acttaagaag	ttacattaaa	149340
atcatctaag	agacagtgtg	acaagaagga	attgtaaaat	gatgttatga	gcacgtgccc	149400
aatgtagtgg	caatcccttg	tgcttcgata	cattgggtggg	agacaaaact	gtacttaaat	149460
tgataaatcc	cttacatgtc	atttttaagga	gcttagactg	actcccatca	tgtagacatc	149520
agagatttct	tttttttttt	tttttttttt	tttttttttt	tttgtgacag	agttttgctc	149580
ttgttgccga	ggctggagtg	caatggcgtg	atctcggctc	accacaacct	ccacctccca	149640
ggttcaagca	attctcctgc	ctcagcctcc	cgagtgcgtg	ggattacagc	catgcaccac	149700
cacgcctggc	taattttgta	tttttagtag	agacgggggt	tctccatgtt	gtggctgggtc	149760
tcgaactcct	gacctcaggt	gatcctcccc	cctcagccac	ccaaagtctt	gaaattacag	149820
gcgtgagcca	ccgcgcccag	cccagagatt	tctaaacaga	gttctaacca	gatgcttttc	149880

58/122

cctgtcagta	gaatgagaat	gaattggagg	tgggagagac	tggcatgagg	gacaccagtc	149940
agccagtggg	attagctggg	aatgttgata	ggagaagaaa	aagattcaaa	gttaggtagt	150000
ggtagcaaga	attagagggg	aggtcggatt	tatgatattg	ccaaggttga	attctaaggt	150060
gaaatttggg	ggcagatttc	atgtgtaaat	tgggaaggta	gattgagttt	ttttaacatg	150120
gggtttctaa	catgtcaata	gagtgactct	gcaggggggc	ctgacgagag	aacagtgcac	150180
gggggtgattc	aacagccagt	tgagccttca	tgcagagcat	ttaacactgt	gactctgtag	150240
actctgggtg	gcagtaaaat	ttcattaaac	caatatttaa	acccttaggt	aataataaaa	150300
attgagggaa	aaggatccag	gttttgtatt	ttttatgaat	tcagttattg	aattaaacag	150360
gaccttgcc	caagaaataa	tctaccaaca	attaacttgt	tttaaagcaa	agttaggaag	150420
tgagcatggt	caaattatta	aataaaaaag	taagctgtgt	atttcattca	tagaaataga	150480
ggctggccta	cttcggatga	ttctcagcat	gtgattacag	atgtgggctt	atacatccta	150540
gggagttaa	gcgtactctg	gcttggtatg	agtagagctc	tttgaaactc	ttctctcacc	150600
cagctagttt	atatagacta	gagaactaga	atgtagcagc	atactctgtc	ttagaagccc	150660
tttataatag	gagctggctg	ggaaggtttg	aaaacataac	aaatgtgttg	gtgtctccca	150720
atgtattgct	agattcctac	ccaagagcat	tatcctgggt	aggggttggg	ttgggtttgt	150780
tttgtttttt	aatgtttgcc	acaaactaac	actagatgtt	agttctttca	tcaagtgagg	150840
agagtagaag	aaaagtccag	aactctgaaa	caccttttca	aaagtttttc	aagccatgat	150900
gtttgcaagt	taaatgctct	gttatgtaag	caatataatc	agttttttat	aatgtaacat	150960
tccttagtgt	tttggggtat	cacacaaaaa	agaatatcca	tatctggaag	caacagcttt	151020
taaataagag	cattgtgggtg	gtgggtggta	tagtgggttt	tttttttttt	tttgagttgg	151080
agtctcgctc	tgttgccag	gttgaggtgc	agtggcacga	tctcagctcg	cttcaacctc	151140
tgtctccag	ttcaagcaat	tcttctgcct	cagcctcctg	agtagctggg	attataggca	151200
cctgctacca	tgcctggctg	atttttatta	tttttagtaga	gacagggttc	accatgttgg	151260
ccaggctggg	cttgaactct	taacctcagg	tgaatcaccc	acctcgccct	cccaaagtg	151320
tgggaattaca	ggcatgaacc	accatggcca	gccaaataag	agcattttta	atgtaaaatt	151380
atgcatgaaa	tgtacattca	attttgtctt	tgtttactag	gatccatgtt	ctcacaagct	151440
atgaagaaat	gggtgcaagg	aaatactgat	gaggtaaatc	ctacctttag	gataaaaaaga	151500
tttctgttta	taagtgccac	cctcatgtaa	gtgaggttta	aaattttcct	tttctttagg	151560
tcccatgttt	aagcagcatg	gcacatttat	gttctcttac	ccagaatgta	ccaagaaagg	151620
gtgggtccctt	cttaacatct	aacaattgcc	tggtagtagc	agtgaaggta	tcttcagtc	151680
gaggctagga	ccactgaagg	atatacatgc	attcaagttt	ccatcagcca	gcaggcatca	151740
gtaatcagtg	tgtagatcaa	aagctcaaat	gtttccttcc	ccactggcag	ttttacttca	151800
agtagtggag	gcttgctttt	ttaatagtta	attaagtaca	ttgagagatg	ggaggtgaaa	151860
aaaggaaaaat	gtttttattt	gaccatctaa	tatgaaagta	gttcgggtgtt	aggtatccag	151920
tagttgacac	tgggaagacag	ggaatgacat	gttaatatct	atagccagag	gggtggcccag	151980
gttttttctg	acatgggaat	gaaattctta	tccaaataag	tagaaattat	gtgcgtaagc	152040
catttgtttaa	gagcactgag	tatgtgcac	tcatccatc	taataaataa	ccattatcac	152100
cagtttaaat	tattttcttt	aggcccagga	agagctagct	tgggaagattg	ctaaaatgat	152160
agttagtgac	attatgcagc	aggctcagta	tgatcaaccg	ttagagaaat	ctacaaaggt	152220
aaggatgact	tcgtttttgt	taaactaaaa	agtattattt	tccaggtgta	aaaataaaaa	152280
agaacataag	gggtttcttt	gcctttgaa	gattaaactc	tgtggggatt	accttcttat	152340
cataagcaac	tagaaaaattg	acaaactaaa	tgaaacaaact	gtttgcatat	attggacaat	152400
gggcaatata	gggaaaccat	ggaaaccaaa	cagagccag	tagtcttgct	gaacgaaaga	152460
gttaaatatc	aaagttcagg	ccaggtgcag	tggctcacgc	ctgtaatccc	agcactttgg	152520
gaggccaagg	cggtggaatc	acttgaggtc	aggagttcaa	gaccagcctg	gccaacatgg	152580
tgaaccctg	tcttagccgg	gtgtgggtgg	aggcacctgt	aatcccaact	atttgggagg	152640
ctgaggcagg	agaatcgctt	gaaccagggg	ggcggagggt	gcagtgaagg	gagatcacac	152700
cactgcactc	cagcctgggc	gacgagcgaa	accccatctt	aaaaaaaaaa	tcaaagttca	152760
gagagctcaa	tttgagtaga	agttgtagga	taaggtagca	gaaaagagga	agctgcccag	152820
aaagaaaagg	gtagagatat	ttagagagat	tccatgggat	ccttggccta	ggagtgatct	152880
gtatatgtgt	gggggtgaaaa	cgcatgtgtc	caggtagaga	accccccaga	aattagtagg	152940
ctgaatgatt	gctggaacat	agggtctaaga	aaagttcatg	gccagaagga	tctggccaga	153000
gtagagagac	ttagtaatac	acaaggcatt	gggtagtgct	ttcacagagg	ttatgcctta	153060
ctactgaaga	taaattagtc	ctagagtaca	agcacctgaa	ccaagtttca	aagcaaattt	153120
ttaaagggtc	aaattaccta	acaactgcac	gccaaaacaa	aggcctaacc	ctctttacag	153180
taacacaaca	aaattcagca	cttcacagtg	taaagttaga	atgtctgacg	tccaggctgg	153240
gcgcagtggc	tcatgcctgt	aatcccagca	ctttggggagg	ccgaggcagg	tagatgacct	153300
gaggtcagga	gttcaagacc	agcctggcta	acatgggtgca	accccgctct	tattaaaaat	153360
acaaaaactt	agccaggcat	ggtggccggc	acctgtgatc	ccggctactt	gggaggctga	153420
ggcaggagaa	ttgcctgaac	ccaggagggt	aaggttgacg	tgagccgaga	tcgcaccact	153480

59/122

gcactctggt	ctgggcaaaa	agagcaaaaac	tcaggctcaa	aaaaaaaaaa	gaatgtctga	153540
cgtcaatcac	aaattaccaa	gcatgacatg	aagttgacct	ataaccagga	gaaaactcaa	153600
tctatagaaa	cagaccacga	tgtgagaaag	atgatgaatt	tagcagacaa	agaccatcaa	153660
gtggctattt	taaatattaa	aaatatgttc	aagtggccag	gtgcagtggc	tcatgcctgt	153720
aatcccagca	ctttgggagg	ccaaggtggg	taggagttca	agaccagctt	ggccaatatg	153780
gtgaaacccc	ttctctacta	aaaatacaaa	aaaattagct	gggcatggtg	gcaggtgcct	153840
atagtcccag	ctatatggga	ggctgaggca	caagaatcac	ttgaaccggg	gaggtggagg	153900
ttgaggttgc	agtaagccga	gatttgtgca	cttgactacc	agcctggaca	acagagtga	153960
actctgtctc	aaaaaaaaaa	aaaaaaaaag	taaagaaaac	aagagtataa	tgagaaaaat	154020
gcaaaatagt	tttaaaagaa	ccaaattggaa	tttcttaaaa	taaaaaatac	cagaaatggg	154080
ggccggggcgt	ggtagctcac	gtctataatc	ccagcacttt	gtgggggctg	aggcaggcag	154140
atcacctgag	atcggtagtt	caaggccagc	ctgaccaaca	tggagaaacc	tcactctctac	154200
taaaaataca	aaattagctg	ggcgtggtgg	cgcattgcct	gtaatcccag	ctacttggga	154260
ggctgaggca	ggagaattgc	ttgaaccggg	gaggcagagg	ttgcggtgag	ctgagattgc	154320
accagtgcac	tccagcttgg	gccacaagag	tgaaactccg	tctcaaaaaa	aaaacaaaaa	154380
aaaacagtag	actcgaagaa	ctagctgagt	ttttctttac	tttaggcagt	aagtgtgacc	154440
ttttgcagggt	gactacttta	gttctctcatg	tcctcattag	tagatcagag	aaattcgaca	154500
ccaaaacccc	aaaagaaaaa	ccccttctaa	tcctcattcc	atgattttat	gaatgcatga	154560
agtcctaggg	ctgcgaagga	atactcattc	tctttatcct	gtgttgatac	ctctctgctt	154620
caacctccaa	ctcgacattt	gcctatagga	tgtacttgga	cattcagcat	aaactacctc	154680
acaccattac	tgaattgctt	catgtgcaca	tgtcccattg	cacaataaccg	gggaccttgt	154740
cttccgtgat	atttgtccgc	agtgtgtgta	ctacaggagg	gagtcagtga	atgtctgcat	154800
gtgtgtcttt	accatccctc	ttgaatatgc	tctagggtta	attcctagaa	gtagaattac	154860
tctattgaaa	attggcaata	tttttcttc	taatatctat	tgccaacatg	ggaaagcaag	154920
tctggatgcc	agtccttggt	atatgccctt	tgggtaagtt	acgtaacctc	tttaagcttc	154980
tgttcactca	tattttaaca	aggaaaatta	caatatttta	cctcacaaaa	ttgtagtctg	155040
cttctggctg	tcttaaactc	tggtatatag	taaacactaa	gtgttggtgt	ccatccttaa	155100
tttgtaataa	taggtcactt	gttagagaaa	tgcaccttac	cattttcttt	tcttttcttt	155160
tttcagttat	gactcaaaac	ttgagataaa	ggaatctgc	ttgtgaaaaa	taagagaact	155220
tttttccctt	gggtggattc	ttcaacacag	ccaatgaaaa	cagcactata	tttctgatct	155280
gtcactgttg	tttccaggag	agaatgggag	acaatcctag	acttccacca	taatgcagtt	155340
acctgtaggc	ataattgatg	cacatgatgt	tcacacagtg	agagtcttaa	agatacaaaa	155400
tggtatttgt	tacattacta	gaaaattatt	agttttccaa	tggcaataac	ccatttatga	155460
gagtgtttta	gcctactgga	atagacaggg	accacatcct	ctgggaagca	gataagcata	155520
gaactgatac	ttgatgcaca	ctcgtagtgg	taactcatcc	ctaatacagca	ttgtaaagca	155580
ggtgccagag	gtgggttgct	ttgtccttcc	aaagcagggtg	agtcagcccc	accgagagcc	155640
aggcagcttt	gagtggcagc	gtgggtgctag	cagcttcagc	ggaacagggt	gagagttaat	155700
tatgcagtct	tcttgacagc	ggcatttaatt	tggaaagaaa	ctgacaagtc	atgggtcaag	155760
tttcagtgcac	ttctctcttc	ctctgatggc	agtatatagt	tttcacattt	taattcctcc	155820
tcttgagatg	cactatactt	aaaaccattc	tctcccctgc	taacagaagg	gtgtgaatct	155880
ggtttacttt	gagcattagg	atttgcccct	ttggaattct	gcactccagt	tacttaactt	155940
tcccttcaga	atacatgtgg	aaagaaagaa	agaaatagcg	atgactccac	ttttgcccct	156000
gtggcacctt	gaacaaagca	gttcttccca	aattatactt	tttttttttt	taaataaggt	156060
gagcaggatg	actggggaga	gagaaacatt	tgactttgac	tgcttcccc	attctttgct	156120
gtgagctgga	aagtgtgcag	ttgggtcgct	ttcttctcct	ttcttttagga	tagtaagaga	156180
ctcactcact	gcacttctgc	tcagttggct	tctgcatcgg	gatcacacag	ccatcagcag	156240
gactgccag	ttgggtgagca	cactccattg	accagtggc	gccagcgctt	cctcaatgca	156300
catgattgag	aggaaagaaa	gttctcttag	atgttactgc	ttttgctcag	actttgcaaa	156360
aaaaaaaaata	tatatatata	tgtataaata	tataattatt	aatcactttt	gtccttgaga	156420
aagtcttgaa	tgaacagaga	atttattcca	ttgcaatatt	tgattgtata	gaggcacact	156480
gtttcatcga	cagaagaagc	aaaaaggctt	tgtgtgaagt	tttggtacta	tgtaccacct	156540
ctgttattct	tttaaagctg	aagtattcat	tgtcttaaac	catattatat	ttaattgtgt	156600
ttgattttaa	aatatatata	tatgaattct	attttaaatt	gtgtcaactt	tctgctttca	156660
gggcattttat	ggctcttctg	ttgaaatata	ttgatcttcc	caaataattt	catttgcttt	156720
ctaaaaaccc	agaacatgag	ccactactgg	actttgcctt	gtgtttgaag	tgtatggcat	156780
aaacccaagg	tttttattag	tcattctatgc	tgtgattaat	tcattttgtt	cttttaacaa	156840
aatattttcca	tccacttcac	attgcttcaa	tctttaacag	aaaagcaata	taaaggttat	156900
agaataaaaat	gtgggttttg	gcaactcttg	ctgcctctgc	atgttttgga	ataacaattt	156960
ctacaagact	ctaggctgtt	taaactagtg	ctttcagtta	agataaattc	taatcatttc	157020
tttgatatata	cattttgtgc	ttctgagcta	gagatgccaa	gtagttgtaa	actgcttata	157080

60/122

aagagaatag	cagcaaat	gagactcg	tacttttt	tgccccac	gctttgag	157140
acagaagc	agtgtggc	gaaattat	gccagatt	atatttgat	taaagtag	157200
ccttgta	attttaa	tggaattg	ttcctcca	attgagcac	caccatgt	157260
caggctct	gcattgtg	cacaaaaa	gattccct	tggaagtt	atgggttc	157320
ataatcag	gaacacc	catcttt	atgttg	cattgacac	aattgttt	157380
aaagaaa	tattagag	aaagtgg	ctttgta	tgatgtgt	tcatcatt	157440
gtaagatt	atgaaagt	aaagcaat	tcagccaa	ccagtgaac	gcaataaa	157500
agggagta	ttttata	tttttct	tggaattc	cattcagta	agcttttc	157560
aatgtaag	gtttacag	ctggagg	gactagtt	gtaggaat	ggagggga	157620
gtcattct	attgtaac	agtacaa	tctttgt	tttattta	tactgagag	157680
taagcac	atgaagt	tgacctct	ccagtga	tgtttggg	cctgcctg	157740
ttcaggag	gggtttat	ttctacac	tgaccttt	tctcgcc	tcctccct	157800
tgccacac	ccagttga	ggacctgg	tgaactc	atccagac	gccccaga	157860
gttctta	ttaagaat	tggggccg	cacggtgg	catgcctg	attgcaac	157920
tttgggag	cgagacag	ggatcact	aggtcagg	ttcgaggc	gcctggcc	157980
catggtga	ccctgtct	actaaaa	cataaatt	ctgggcat	tgccgcag	158040
ctgtaat	agctacgt	gtggctga	cagggga	gcttgaac	ggaggcgg	158100
gttgtgca	gagccgag	cgtgtcact	cattccag	tggtgacag	agggagac	158160
tgtctccaa	aataaaaa	agaaaaag	ttttggg	ggtgcagt	ctcacgc	158220
taattacag	attttgga	gccccag	ggcagat	ttgaggac	gagttcg	158280
ccagcctg	caacatgg	aaactcc	tctactaa	agacaaa	tagccag	158340
tggtgatg	cacctata	cctagct	cgggagg	gggcagg	atcacttg	158400
cccagga	agagattg	gtgagcca	atcacat	tgactccag	cctgggca	158460
agagcaag	tctgtctc	aaaaaaag	atttggc	gcgcagtg	tcacgc	158520
aatcccag	ctttggg	ccaaggc	cagatcac	ggtcagg	tcgagatt	158580
cctggcta	atggtgaa	cctgtct	ctaaaaa	aaaacatt	ccgggtgt	158640
tggtgggc	ctgtagtc	agctact	gaggctg	cagagga	atgtgaac	158700
aggaggcg	gcttgcag	agccaag	gtgccact	actacagt	gggcgac	158760
gtgagact	gtctcaaaa	aaaaaaga	tttggcc	tgcggtg	catgcctg	158820
gtcccagc	tttgggag	caaagtgg	ggattac	aggtcagg	ttcaagac	158880
gtccggcc	tatggcga	ccctgtct	tactaaaa	aatacaaa	ttagccag	158940
gtggtggc	gcacctgg	aggctgag	agggagaa	gcttgaac	gggaggca	159000
ggttgcag	gtgccag	gtgccact	actccag	aagactct	ctcaaaaa	159060
aaaaaaaa	aattttgc	ggggaagg	agatact	caccatct	aatggtgt	159120
ggatgtgg	cttacaaa	caggagcc	cactgcat	acaaacag	gcattgtg	159180
ctgagatg	aggtacct	ataaccct	agacatct	ggtttctg	tctattct	159240
catccttg	ttggacta	ttaatctg	agttatct	ataatgat	ttgatttt	159300
ttttttga	tggagtct	ctctgttg	ccaggctg	gtgcaatg	acgatctc	159360
ctcaccac	cctccact	ccaggttc	gtgattct	tgccctca	tcctgag	159420
ctgggatt	aggcatgc	caccacac	ggctaatt	gtattttt	tagagacg	159480
gtttctcc	gttggtc	ctggtctc	actccca	tcaggtga	accctgtc	159540
ggcctccca	agtgtcgg	ttacaggc	aagccatg	accgggtc	ttttttg	159600
ttttgaa	agtctga	gagttttt	aattacgt	aaggagtt	gctaaaa	159660
tgccata	ccctaagt	taatgatt	gtattctc	catgtctg	aagtactg	159720
gatttctg	gaataatt	tctttagt	acttcact	agtcgtc	tgtattct	159780
caaaatg	tcctaacc	atggagct	aagacac	ttgttttt	aacaagc	159840
tactaggc	caggaagg	agaagtcc	ggcttgtg	atgatcac	ttagaact	159900
ggcctggg	agtgcctt	catgcttc	agatcctt	aaagaata	gaagatt	159960
accgcttta	gcaattgt	taaacc	aatagaa	tttttggt	gagtag	160020
agaagttg	cgggagag	aattttta	aaattgt	atacctg	attctata	160080
ctaggcaa	tctctggc	tgtaaa	ctcaagg	caacttgg	ggccacac	160140
aatagtt	cactgagg	ctctccgg	gaacattg	cactagag	agccctct	160200
cttgggc	actgggc	gtgcagag	ggagcgg	tactgtgg	tctgggc	160260
tgagatgg	cagtgatg	caataa	cactggag	agcagtgt	gtaaagg	160320
tgagggc	catgttc	gagggtt	gcccact	ttgcttgg	cacaggag	160380
tggtattcc	actttatg	actttatg	tctcagt	ctcagtc	gttcctct	160440
aactgcct	tctggcac	tcagaatg	acagaact	cctggatg	ttcagccc	160500
tgccataa	tgacagt	tctccttc	cacccacc	ctcatacc	tgaagcac	160560
gtcagact	cccagtc	gggcaagg	cctagag	gcttagtt	agcttgaa	160620
gagctggg	ttaccaag	gcaaatga	gacgagg	gcaaca	tgccatttc	160680

61/122

ccagcttccag	ctgactcctg	tatatgtact	gtgccttccag	actcatccgt	aagtgacccc	160740
aggctggcct	ctccacatc	acagtaagaa	ttccacacac	catacaactt	ggaaagaggg	160800
tccagctgaa	ggaagcccca	cacttcttcc	aagtttttct	tagtcttctc	ttcttgga	160860
agagtacctt	ttgtttcttc	taattatgta	actattgggt	tagtaaatat	tcacccattc	160920
agtcaccctg	taagtggcag	gcactgttta	cagggacaca	ggaaggaata	aaaacttgca	160980
ggcaccttgg	agcttgcat	ctattgaaga	ggtaatggaa	gttgggtag	cagctaaact	161040
atgctgggtat	tggccaggcg	cagtggctca	cacctgtaat	cccagcactt	tggaggccaa	161100
ggtgggcaga	tcatgaagtc	aggagatcga	gaccatcctg	gctaaccatgg	tgaacccccg	161160
tctctactaa	aagtaaaaaa	aaaaattagc	caggtgtggg	ggcgggcgcc	tgtagtccca	161220
gctacttggg	aggctgaggc	aggagaatgg	tgtgaaccca	ggaggcgaag	attgcagtga	161280
gccgagatgg	caccactgca	ctccagcctg	ggtgacagag	cgagactctg	tctcagaaaa	161340
aaaaaatatg	ctggtagttt	tgattcaaga	tggcctttgg	agcccatgat	ttaggtctcg	161400
taccacccaa	ggtctactgg	aaaacatcag	gctctcctgc	tatagaccca	tagggagagc	161460
tgcagccgag	agggggagct	gaagagaagt	gccccctctg	tgtcctgtca	gcctcatcct	161520
tccgcaagga	ccagttgtct	tgccactcca	ttcacttgct	gcaagactgg	aggtttttcc	161580
tcaggtgttg	agcactgggt	ttacaagatg	tcagcatctt	gatgcctgag	accatcaagg	161640
caagtctctg	aacagggtct	accttagagt	aaggcttaga	agaggccgta	aagtcagtct	161700
cagctccgtg	gctctgcaga	gctttgggac	atgtgaattc	ttaaaaacaa	gactattgta	161760
cagttactat	atgcatgcag	tataaaatta	taaccttgga	aaatcctagc	tagctgttga	161820
gctaattcca	taaagtaatc	agctcctgag	ttctgcagtg	gtaataataa	tcagcataat	161880
gagtaaacac	tgtgtgtgcc	aggcagcgtc	tcatttgatc	cttgtgataa	tcttgtaagt	161940
actgatttcc	tcccttcttt	aaacaaagtt	tttttttttt	tttttagagag	ggtctcacta	162000
tgttgcccg	gctagtcttg	aattc				162025

&lt;210&gt; 36

&lt;211&gt; 162025

&lt;212&gt; DNA

&lt;213&gt; Homo Sapien

&lt;220&gt;

&lt;221&gt; mutation

&lt;222&gt; 156,277

&lt;223&gt; Nucleotide Base Change: T to C

&lt;400&gt; 36

gaattcctat	ttcaaaagaa	acaaatgggc	caagtatggt	ggctcatacc	tgtaatccca	60
gcactttggg	aggccgaggt	gagtgggtca	cttgaggtca	ggagttccag	gccagtcctg	120
ccaacatggg	gaaacactgt	ctctactaaa	aatacaaaaa	ttagccgggg	gtggtggcgg	180
gcacctgtaa	tcccagctac	tcaggaggct	gaggcaggag	aattgcttga	acctggggaga	240
tggaggttgc	agttagccga	gatcgcgcca	ctgctctcca	gcctgggtgg	cagagtggaga	300
ctctgtctca	aaaagaaaca	aagaaataaa	tgaacaattt	ttgttcacat	atatttcaca	360
aatttgaaat	gttaaaggta	ttatgggtcac	tgatctctg	tttcattctt	tataaatca	420
ttaagtttga	aatgtatact	tgactacta	acacagtagt	taatcttagt	cctacaagtt	480
actgctttta	cacaatatat	tttcgtata	tgtatgcact	ggtgtttatg	tacgtgttta	540
tgtttatata	tgttaaaatt	agcagtttcc	atctttttct	attttgtacc	atcacatcag	600
ttcagaagga	ttgacagagc	aaaatgattt	gatgaagtat	aaaagtcaca	tggtgagtg	660
cataaatata	actctgaaca	attaggaggc	tcactattga	ctggaactaa	actgcaagcc	720
agaaagacac	atatcctata	tgtcaagaga	tgtaccaccc	aggcagttaa	agaaggggaag	780
tacacataga	aagcacatag	gtgaataatt	aaaaaattgg	aatttatcag	acactggatt	840
catttgctcc	taaagtcaga	gtcctctatt	gtttttttgt	ttttgtgggt	ttctttttta	900
atttttttat	tttttgtaga	gtcggagtct	cactgtgtta	cccgggctgg	tctagaactc	960
ctggcctcaa	acaaacctcc	tgccctcagc	tcccaaagca	ttgggattac	agacatgagc	1020
cactgagccc	agcccagacg	ctttagcatt	tatgaagctt	ctgaaatagt	tgtagaaacc	1080
gcataagctt	tccatgtcac	tttcaaagtt	tgatgggtct	tttagtaaac	caaccaagtt	1140
attcctcaag	ggcaaaaata	cattttctcag	tgcaaaactg	atgcacttca	ttacaaaaag	1200
gaaaagacca	caactataga	ggcgtcattg	aaagctgcac	tcttcagagg	ccaaaaaaaa	1260
aggtacaaac	acataactaat	ggaacattct	ttagaagagc	cccaaagtta	atgataaaca	1320
ttttcatcaa	agagaaaaga	gaacaagggt	ttagcaaaat	cctctatcaa	ataacactaa	1380
acatcaagga	acatcaatgg	catgccatgt	ggaagaggaa	gtgctagctc	atgtacaaac	1440
cagtagataa	tttcaacttg	ctgccgaatg	aaacctcttt	gcaagggtatg	aatcagcact	1500

62/122

tctcatgttt	gttttgcctt	gttttgtttt	gtttttagag	acaggccctt	gctctgtcac	1560
acaggctgga	gtgcagtggc	acgatcagag	ctcactgcaa	cctgaaactc	ctgggctcaa	1620
gggttctctc	tgccttagcc	tccaagtag	ctgggactac	aggcccacca	tgcccagcta	1680
atttttttaa	tttttctatag	agatgggatc	tcactagcac	ctttcatgtt	tgatgttcat	1740
atacaacgac	caaggtacaa	tgtggaaaag	gggtctcaggg	atctaaagtg	aaggaggacc	1800
agaaagaaaa	ggggttgcta	catagagtag	aagaagtgc	acttcatgcc	agtctacaac	1860
actgctgttt	tcttcagagc	agagttgatg	atctaaatca	ggggtcccca	acccccagtt	1920
catagcctgt	taggaaccgg	gccacacagc	aggaggtgag	caataggcaa	gcgagcatta	1980
ccacctgggc	ttcacctccc	gtcagatcag	tgatgtcatt	agattctcat	aggaccatga	2040
acctatttgt	gaactgagca	tgcaagggat	gtaggttttc	cgctctttat	gagactctaa	2100
tgcgggaaga	tctgtcactg	tcttccatca	ccctgagatg	ggaacatcta	gttgaggaa	2160
aacaaacctca	gggctcccat	tgattctata	ttacagttag	ttgtatcatt	atttcattct	2220
atattacaat	gtaataataa	tagaaataaa	ggcacaatag	gccaggcggtg	gtggctcaca	2280
cctgtantcc	cagcacttcg	ggaggccaag	gcaggcggtg	cacgaggcca	ggagatcgag	2340
acctctctgg	ctaaaacggg	gaaaccccg	ctactaaaaa	ttcaaaaaaa	aattagccgg	2400
gtgtgttgtt	gggcacctgt	agtcccagct	actcgagagg	ctgaggcagg	agaatgggtg	2460
gaactgttga	ggcagagctt	gaggtaaagg	gagatcacgc	cactgcactc	cagcctgggc	2520
gacaaatcga	tactctgtct	caaaaaaaaa	aaaaaaaaaa	aaagaaataa	agtgaacaat	2580
aaatgtatgt	tggctgaatc	attccaaaaa	aatcccccca	ccccagttca	cggaaaaaatt	2640
ctcccaaaaa	accagtcctc	ggtgccaaaa	aggttgggga	ccgctaattct	aaataatcta	2700
atcttcttct	aatgctaaaa	aatgaataaa	ctttttttta	aatacacggg	ctcactttgt	2760
tgccttctct	ggagtacggg	ggcatgatca	cagctcactg	tagcctcaat	caccagggcc	2820
ccagcttctc	tcccacctaa	acttctctgag	tagctgggac	tacaggcacg	caccacctatg	2880
ctcaaaacct	ttttaaattt	tttatagaga	tgggggtctc	accatgttgc	ccagactggg	2940
cttgggtctc	gggctcaagt	gatacctcct	caaactcctg	gactcaagtg	atcctccttc	3000
tttttaattc	caaagtgtct	ggattacaag	catgagccac	tgtaccacag	tggataaaca	3060
gttcccaagt	cactacagtc	atggacaatc	aggcttttca	acatgcagta	tggacagtga	3120
aaggatattt	atctgtttcc	atactgaaat	acatgtgata	ctaaggagaa	aggtgtctgc	3180
ataaggtctg	aaaatgaaga	atatttaaaa	tgaggaaaaa	actgtttctt	catgactttg	3240
gtaatcattt	ttaaagaccat	ttctgtgatc	tcaggtgatt	cactcaagta	gtatatattca	3300
tatgatacct	tctggaacag	cctgaatctt	aaccaaaaata	ccatgatttt	ttaatgctgt	3360
acttccccga	tgatgatatg	accaaactgc	aatgtaggca	gctaaatctc	cacgagtttg	3420
attggcatat	gagttgacag	ttttcttcac	aaattaaaga	aatatatatt	ttgatacatg	3480
gaatcaaatg	ttaaaaacta	cactgaaatg	ctgcaaaatg	atataaagaa	acattttcca	3540
acacttgjac	caatcaaaga	gtggattagg	aatctactca	ccattatcaa	ctaaatagaa	3600
agggtggatt	tgggtgtggt	ggctcacatc	tgtaatctca	gcactttggg	aggccaaggc	3660
ctctacaaaa	cttgaggcca	cttgagctca	accagctcga	gcaacatagc	aaaactctgt	3720
ctctggaagc	aaaaaataaa	attaaccagg	catggtggca	gatgcttgta	atcccagcta	3780
tggctcatgt	tgaagttaga	ggactgcttg	agcccaggag	atcaagactg	cagttagaccg	3840
aacaaaaaac	gcgccacagc	ctgagtgaac	gagagagacc	ctgtctcaaa	aacaaaaaca	3900
gatggagtct	acttaacctt	cctgtttttt	gctgttgttg	ttgttgtttg	tttgttttga	3960
agctctgcct	cactctgttg	cccaggctgg	agtgcagtgg	cgtgatcttg	gctcactgca	4020
taggcgcccc	cccggtttca	cgccattctc	ctgcctcagc	ctcccagata	gctgggacta	4080
cgtgttagcc	ccaccacgcc	cggctacttt	tttgcathtt	tagtagagat	ggggtttcac	4140
aagtgtctgg	aggatgggtc	tgatctcctg	acctcgtgat	ccacctgcct	cggcctccca	4200
gatattgctg	attacaggca	tgagccaccc	cacccggcca	acctttctgt	tttttagttt	4260
tcactagaat	ttaaactcag	agctgaaaga	atgtgaaaag	tggccttcag	taaaaaattt	4320
caaaaaacact	ctctacatcc	atatttaate	tgaatgcata	tccagattga	tcagtttagag	4380
tggaaacaga	catcatcatt	cctgatgacc	tctaattctg	gtttcggctt	tctatttcaa	4440
ctatcaaaaga	ataaggaaag	aaatgggaag	gctctggaaa	tttgtcctgg	gctatagata	4500
ttttaattat	tcaccaacaa	taagatctct	ccataaaata	taaaacaagt	ataatttaatt	4560
atactattga	ttttttctct	tcagaggatt	ttattttcaag	ataaaaacata	acttctaccc	4620
cacagcaatg	ttccaaagg	tagaaaaagt	gtttttcctc	atcttatcct	tcaaagaggt	4680
aatcattgaa	caaacatcta	taaaatgcct	ctgcataatt	gtcagaagct	atagtccaga	4740
acaacttact	aatgtctttc	cattttaaac	ttaggtgagg	tgtcttagga	aacctctatg	4800
atctcatttt	ctattttatt	ggaggtaaac	tcccagactc	tcccagggtc	tcctgtattg	4860
ttctgcacat	ttaggtctcc	taatcccttg	aagcacatc	gaaaaagccc	tggatctctt	4920
agcggcctcg	atcatcgcg	aattcattcg	gcttccagca	agctgacact	ccatgatata	4980
gctgggcttc	cccttctccg	gacgccagtc	cttgctgcgg	ttagcttagga	tgaggggttt	5040
	agtgcaggct	tctgcggggt	cccaagccgc	accagggtggc	ctcacaggct	5100

63/122

ggatgtcacc	atttgcacact	gagctcctgg	caggetgtac	caattttttta	attattttaat	5160
atttattttt	aaaattatgg	tgaatatattt	ggtattctgc	tctaaaatag	gcccataaat	5220
gcacagcaga	tatctcttgg	aaccacacagc	tttccactgg	aagaactaag	tattttttctt	5280
ttaaagatgc	tactaagtct	ctgaaaagtc	cagatcctct	acctctttcc	atcccaaact	5340
aagacttggg	atttatgaga	gatctagcta	acagaaatcc	cagacacatc	attgggtctt	5400
cccagagtgc	agtcctccta	aagaggctca	gccctaagca	ggcccttgca	ccaggagggt	5460
gggtctgaga	cccacatagc	acttcccaag	gtgcatgctc	cagagaggca	ctgaaacagc	5520
tgagcacaag	cctgcaagcc	tggagaactc	tcacagtcag	aacggagggg	gcccagtggt	5580
actaacataa	agagaaaagg	gaacacagag	aaatggatgg	caccaacaac	cagcaaagcc	5640
ttcatggcca	atgaaagcat	cagtgaagg	gccagaacct	tcaccccaa	agactcttca	5700
ctgcctttag	tgaaaaacaa	tggctagaga	gtgaagtatt	gatcatgtat	agagaggtaa	5760
agttacattt	ttatatcttg	actctgctaa	tgtgaaattc	cctatctgct	agactaaaag	5820
tttcagacac	cctgttcaaa	tatcccatata	gttgctagag	acttaaaatg	aacagaacgc	5880
acattgtcag	gatgactatt	accaaaaaat	caaaagacag	caagtattgg	tgaggatgta	5940
gagaaaactgg	aacttttggg	cactgtttat	gagaatgtaa	aatggagcag	ctgctgtgga	6000
aaagagtatg	cagggttcctc	aaagagtaaa	accaagatgt	ggaaacaact	aaatgcccat	6060
cagtggatga	aggggttagac	aatatgtgg	atatacatat	catggagtac	tattcagcct	6120
ctaaaaaaa	aaaaggaaat	tctataacat	gcaacagcat	ggatgaatct	tgaggacatt	6180
ttgctaataa	aataaggcag	tcatagaaag	acaaatactg	cacgactcca	cttatatgag	6240
ataccaaaaa	tagacaaatt	catagaatca	aagagtacaa	tggaggttac	ctggagctgc	6300
agggcgggaa	acgaggagtt	actaatcaac	gaacataacg	ttgcagttaa	gtaagatgaa	6360
taagctctca	agatcagctg	tacaacactg	taccatagagt	caacaataat	gtattgtaca	6420
cttaaaaaat	tgtaagggtt	agattaacaa	atgtagtaga	tccacaaatg	tggttaagtg	6480
ttcttaaccac	agtaaaaata	aaaaagaata	tcaagcccag	gagttcgaga	ctagcctggg	6540
taacatggtg	aaaccctgtc	tctacagaaa	atacaaaaat	tagccagctg	tggaggtgca	6600
ctcctagggg	ggctgaggtg	ggaggcttgc	ttgagcccag	gaggtcaagg	ctgcagttag	6660
ccatgattgc	accactgtac	tccagcccag	atgacagagc	aagacaccac	cccccccaaa	6720
aaaagaaaaa	gaatttcaaa	cattttaaaa	gatcagatac	gcaagaacaa	caacaaaaaa	6780
gagatgaaca	gagcatcgac	cctcatctag	tgggattcct	ggtctaactg	aaaaacagac	6840
attgagagac	aaacaatgac	agtgatgtga	tcacagcaat	tacacaggta	tcccctgggg	6900
actgcagaag	aaaggaggaa	tgccctaactt	tcagaaaata	gagaaagcgt	caaacagttg	6960
gtgaaagcct	tccaaaaacta	gagagaactg	cacacaccaa	atcacagaaa	gaagaaaagc	7020
cgtgggagat	tctgggaccc	accggctatt	tttgatggct	gaacaccctg	ctgcaggaga	7080
gacaggagct	ggaaagcatg	gtgggatgaa	acctcaaaac	gctttgcctg	cattgcttaa	7140
gatgactggg	cttgattaac	tctagtcaat	ggggacaatt	caatcaaaga	agaaagatgc	7200
tcaaattcac	attttagaat	gattttttat	ggcagtatgg	ggaatagatt	aaaagagagt	7260
gaagctggag	gcaagaaact	tgtaagagg	caactgaaac	agtctagatg	ataaataata	7320
aactgacaga	gtgactagaa	aaatcagaac	aggctgaatc	aacagatacc	tagatgaaaa	7380
taacaggact	tgatcaccag	ttgtatcttg	gagaggaagg	agttgtttcc	ttgctttccc	7440
tacgactggg	aatacggga	gtttgcccgtg	tgtattgggt	atatactgg	gtgtagccaa	7500
tcactgacaa	ccatttagca	gcttaaaaaca	caaaggctta	tctcccagtt	tctgtgggac	7560
aggaatctaa	gataggctta	gctggctgg	tctggctcag	agtttctcaa	gaggttgcaa	7620
tcaagatgtc	agctgggggt	gcatcatctg	aaggctcaac	tggggccgga	gggtccactt	7680
ccaaggagtt	cactcacctg	cctgacaagg	cagtgtgggt	tgttggcagg	agatctcaat	7740
tcattgccaa	gtgagcctct	ctatagcatt	gctggaacat	cctccccatc	tggcagttgg	7800
cttctctcag	catgagtgat	ctgagagaga	gagcaaggag	gaagccacag	tggtcttctt	7860
actcctactc	ctaacactat	ggacctaact	ctaacactct	cacttctgcc	ttattccatt	7920
agttagaaa	ggaactaagc	tccacctctt	gaaataagaa	gtgtcaaaga	atttgtggat	7980
atatttaaaa	atcatcacac	tgtggaagtg	gataggggg	tcaattaatg	ctgaacttga	8040
aatgcctgag	acattcaaat	gtccaacagg	caatgaacat	acccatagat	ggtcatgact	8100
ttagacaaga	tagaggaaga	tcacagaatt	aaggaggaat	tgaaaggtaa	aagaagtggg	8160
gtcagattcc	ccctgaaaag	tgagccatga	aaggaaactt	aactattgag	ttagaggtca	8220
gagtaggaaa	tttcgggtgga	attctttttt	aaagaaagga	accatataag	catgttttga	8280
ggtagagggg	gaataaatca	gtagacaggg	agaggtaaaa	aacataaatg	ataggggata	8340
ggtgacaaag	gtcttggcag	aatcccttac	ccattgactt	ggggccaaga	gagggacact	8400
tctttgtttg	agggataaag	aaaataagaa	agaatgggtg	ctatttagtg	tggtcctgtc	8460
tctagggcaa	acgcataaggt	aacaaactgt	gtgtgttagg	aatatagatg	tgacctcaca	8520
ttgagattct	cacctcaaat	ccattttgtt	gttacctgta	ccttcctacc	ttctcttttt	8580
gctacatgca	gactgctgtt	ttgtcttctt	ggcctgttcc	aggtttcagc	attctggcat	8640
atctgctacc	ctgttcccaa	acctctctag	agtcctatgt	ccttccttgg	atagtgtttg	8700

64/122

attggggccac	gtatctaaga	agtgatgcct	tcagtttaggc	ctgagaacct	cctctatgga	8760
aatctccatc	agtgaccctg	acagacttgg	tatcttggag	atgtcactgc	ccccagcctg	8820
tgggtctagga	gaatctcagc	ctgggcctct	agtagtatgg	ataaggcggt	aagggtatctt	8880
tgaaccagag	tctgtcatat	tcctcaatgt	gggacagata	aaacagtggt	agtgtctggtg	8940
tttctgagct	agaactctgg	tttttgggtc	agattctttg	atgtatgacc	tttcagaggt	9000
attaaaattt	gttctaatac	aatgttcaat	acaaatgtag	ttccttttct	gttaggacct	9060
caacaaaaca	tgaccaactg	tagatgaaca	ttaaactatg	acaattcatg	gaaatgaata	9120
cagtaatacc	tgcggttccc	ccatttttagc	agtcactatg	gtgacatttg	gcacaaatgg	9180
ctattttaagg	gtgcttttgt	taaaacctac	catcttacta	ggcacatgat	attgaaacta	9240
atgaaataat	ggagaaaact	cttaaaaaact	tttaatgaat	aaagtgatga	agtgataata	9300
tttttagctgc	tatttataaa	gtgactatta	cagggtcaaac	attcttctag	ggtttttttg	9360
ttgaagtgtg	cacatttaat	ccttaataac	ccactatgag	tcagggtattc	ttctctcccc	9420
tttggacagt	tgggggaaatg	gggggtcagag	agggttaggta	atttgctcag	ggccacacaa	9480
cctgcagtga	gaaaactctga	gatttgtaca	ggaacgtatc	aaactctgaa	gtccatgctt	9540
ctattttccc	agctgtgcctt	tctaataaaa	ggttaactaat	gctactggat	gctgccccca	9600
aagtgaagtca	ctttcaccccc	accctacttg	atttttctcca	taaaactaat	cacatcctga	9660
caacttattt	attgctgatc	tcccccacta	gattataaac	tcaataaaaag	caagatcctt	9720
gtctgtctgaa	tatcagtacc	taaaacgctg	tctagcacag	agcaagtaat	taatatttgt	9780
tgaatgaaca	aataaaggaa	aaaaattcaa	aggaagaaaa	agccctaaaa	cagatgttta	9840
cctaaacata	catttttaaaa	gaaagcatat	aacaaattca	ggacagaatt	taaatttgat	9900
tttttaaga	aataaccaag	tgctagctgg	gcacagtggt	tcacacctgt	aatcctagca	9960
ctctgggagg	ccgaggcagg	cagatcactt	gagggtcaaga	gttcaagacc	agcctggcca	10020
acatgggtgaa	acctgtctct	actaaaaata	cagaaattat	ccaggcatgg	tggcagggtcc	10080
ctgtaacccc	agctactcag	gaggctgagt	caggagaatt	gcttgaaccc	aggaggcaga	10140
gggtgcagtg	ggccaagatt	gcaccactgc	actccagcct	gagtaacaaa	gcaagactct	10200
gtctgaagga	gaaggaaaga	aagaaggaaa	gaaggaaaga	aggaaagaag	gaaagaagga	10260
aagaaagaaa	gaaagaaaga	aagaaagaaa	gaaagaaaga	aagaaagaaa	gaaagaaaga	10320
aagaaagaaa	aagaaagaaa	gaaagaaaga	accaagtgtc	tatttgggac	ctactatgct	10380
atgtttttcc	atgcacgcta	ttttcagtaa	agcagttagc	aaacttgcaa	gatacataaca	10440
acaaatatat	gcttctataa	ctctaaaatt	gtgcttttaag	aaagtccctct	ttaccagctc	10500
atgtatgcac	tagttttcta	agagttaacta	gtaacttttt	ccctggagaa	tatccacagc	10560
cagttttatt	aaccaaagga	ggatgcttac	taacatgaag	ttatcaaagt	tgagcctaag	10620
ttgggcccagt	tcagtgttaat	atactccaga	acaaaaacca	tcctactgtc	ctctgcaaat	10680
tttacctgaa	aattcatttt	ccacattacc	aaggagccag	ggtaggagaa	tatagaaaga	10740
ccacccaaga	atccttactt	ctttcagcaa	aatcaattca	aagtaggtaa	ctaaacacat	10800
gccctaacaa	tgaatagcag	attgtgctca	gaagaatgat	ctacaacatc	ttactgtgaa	10860
ggaactactg	aaatatctca	ataagacttc	tctccaaaat	gattttattg	aatttgcatt	10920
ttaaaaaata	ttttaagcct	aaatttttaa	aggtttgata	ttggtacatg	aatagacaaa	10980
cagacatgga	ctagaccaag	aattaggttc	aaacatatac	aggaatttaa	tatacgataa	11040
atctagtatt	ccaaaggaac	caacaaatgg	tgttcagaca	gcaggatagg	catcaggaaa	11100
aacacagttg	ggcacccctac	cttactccta	acaccaggag	taactgaagg	agcaccaaat	11160
atttatttat	tttaattata	gttttaagtt	ctagggtacg	tgtgcacaac	atgcagggtt	11220
attacatagg	tatacatgtg	ccatgttggt	gaggagcacc	aaatatattaa	aagaaaaaaa	11280
ttggccaggg	gcggtgggtc	acacctgtaa	tcccagcact	ttgggaggcc	aagggtgggca	11340
gatcacctga	ggtcgggagt	tcgagaccag	cctgagcaac	atggagaaac	cccatctcta	11400
ctaaaaatac	aaaattagcc	aggcatgggt	gcacatgcct	gtaatcccag	ctacttgga	11460
ggctgaggca	ggagaatagc	tttaattctgg	gaggcacagg	ttgcggtgag	ctgagatatt	11520
gcaactccagc	ctgggcaaca	agagcaaaac	ttcaactcaa	aaaaattaat	aaataaataa	11580
aaataaagaa	agaaaagaaa	aaaatgaaaa	tagtataatt	agcagaagaa	aacaccgtag	11640
aatcctcgga	ctcttaggat	ggggaatgcc	tataatataa	aaaccctgaa	gttataaaaag	11700
agaaaatcac	ctacatacaa	accaaactct	tctacatgcc	taaaacatag	cacaaacaca	11760
gctaaataat	catagctgaa	tgaactggga	aaacaaaact	tgactcatat	ccagacagag	11820
tttaattttcc	tacacataaa	gagtacctat	ataaacccaa	caaaaaaacc	accactaacc	11880
caaaaataaaa	atgtgacagg	taatgaacag	gtagttcaca	gagaatacaa	atgggtcttc	11940
ggcacataag	tgacttttac	tgacttttac	ttattttatt	tttgagagac	aggggtctcac	12000
gatgttgccc	agggttaggt	caaactcctg	ggctcaaagt	atagtaccag	gactacaggt	12060
gtgccccacc	gcacctggct	cctcaaccac	ctgtattaac	aggaaatgca	aaataaaaact	12120
ttcaaatcta	ttttacctat	tagaatggca	aaaatttgaa	aaacttcaaa	catcatcatg	12180
ttgggtgagaa	tgtgaggaga	ctggcactct	catttttttgc	tgatagcata	tatatactga	12240
tggcttctat	ggaaagcaat	ctggcagcgt	ctatcaaagt	tacaagtgca	tatatccttt	12300

65/122

gacaaagcaa	ttccactcta	ggaatgtgtt	ctatatgggt	gtgcttctctg	gggctgggaa	12360
ctgggagcta	agggacaggg	gcagaagata	atcttctttt	ccctccttcc	ccgttaaaca	12420
tggtgaattt	tatatactgt	aataatattat	ttttcacaaa	agataatttt	taagcgatat	12480
gtctgggaat	tttttttttt	cttttctgag	acaggggtctc	actctgtcat	ccaggctgga	12540
atgccatggg	atgatctcag	ctgactgcag	cctcgacctc	ctgggttcaa	gcaatcctcc	12600
cacctcagcc	tcctgagtag	ctgggactac	aggcacgtgc	catcatgcta	atttttgtat	12660
atacagggtc	tcactatggt	gcccaggcta	atgtcaaaact	cctaggctca	agcaatccac	12720
ccacctcagg	ctccaaagtg	ctgggattac	aggcgtgagc	caccgcgcct	ggccctggga	12780
attcttacaa	aagaaaaaat	atctactctc	cccttctatt	aaagtcaaaa	cagagaagga	12840
aattcaacct	ataatgaaag	tagagaaggg	cctcaaccct	gagcaacaaa	cacaaaggct	12900
attctgaga	caggaatttg	ctgaacaaaa	tcgaggggaag	atgacaagaa	tcaagactca	12960
cttctcggct	gggcgagtg	gctcacacct	gtaatcccag	cactttggga	ggccgaggcg	13020
gacagatcac	gaggtcagga	gattgagacc	atactggcta	acacagtga	acccagtctc	13080
tactaaaaat	acaaaaaatt	agccgggctg	gggtggcagg	gcctgtagtc	ccagctactt	13140
gggaagctga	ggcaggagaa	tggcgtgaac	ccaggaagcg	gagcttgag	tgagccgaga	13200
tcacgtcact	gcactccagc	ctgggtgaca	gagcaagact	ctgtctcaaa	aaaaaaaaaa	13260
aagactcatt	tctctagatc	ttgagccgta	ttcaaattta	tctcagctta	gtgagagggt	13320
aaagcaagg	atatccttcc	ctgtgggccc	tgctccttac	tgaagggaag	taacggatga	13380
gtcagagaca	ccaatggaga	aaagcactaa	caccattatc	tgatgaacat	tacgtgaaga	13440
aggttagaa	gtgaagtgg	attgctgaag	aagtcagtga	aagcggacat	tcatttgggg	13500
aaatggata	taggaaatcc	ataaaagtga	ttaaaaagat	gttagaggct	gaggcgggg	13560
gaccacaggg	tcaggagatc	gagaccatcc	tggctaacac	ggtgaaaccc	catctctact	13620
aaaaatcaaa	aaaattagcc	aggcgtgggt	gcaggcacct	gtagtcccaa	ctactcggga	13680
gactgagga	ggagaatggc	atgaacctgg	gagacggagc	ttgcagtga	ccgagatcac	13740
gccactgcac	tccagcctgg	gtgacagagt	gagactccat	ctcaaaaaaa	aaagttagat	13800
acgagagata	aagatccaac	agacacacaa	ctgctaattc	tgaacagaac	aaaacaaatg	13860
gcacaggaag	agaaaaattta	agatataaca	ccggaaaaact	ttcctgaaat	tgagtaactg	13920
aatctatagc	ttgaaaagggt	ttagcatatg	ccaagaaaaa	tcagtagagt	ccaaccagca	13980
caagacacat	ctagcaaggc	tggtgattct	accaaacacag	agaaagaagt	gggtgaccca	14040
taatgcggaa	aaaggcagac	catctgcagt	cttctccaga	acactggagt	ctgaagacaa	14100
aagaatgctg	cctactgagc	cagaagggag	agaaagtga	ccaacacatc	tttaccaagt	14160
tagaatgtca	cgcattattt	aaaggctgca	aaagccatga	aagacatgaa	agaacacaag	14220
catttacaa	atgaaagaac	acaagcattc	gcatactcaa	gaatccttaa	gaaaaatgta	14280
gtcctaatec	agcccactga	aagttaaattg	tacttaattg	gctcattaat	gggaacttca	14340
tagcttcaaa	tcagtctggg	cccattctacc	aacatctctc	gccccggctt	cctgcaatag	14400
tcagcacctt	tccctcctcc	cagtcttgte	ccctggagtc	tgctctcagc	atagcagagt	14460
gaccacacta	acacccaagt	cagagccctc	cagtgcgcac	tggtctacaa	agcccttccc	14520
acccccacac	ccacgtgccc	tccggatcct	tgtgacgtgt	ctcctgcata	ccctagcagc	14580
cctggcctcc	tcactgcccc	tcctgtacat	caggaaggcg	actccttgag	tcttggtctt	14640
ggccgcctcc	tccacctgca	gtgagttaac	tcccttacct	actctaggtc	attgctcaaa	14700
tgctcagcat	tcaatggggc	cctccctgac	taccctattt	aaattctaca	tactcccctt	14760
gaccccatgg	acctcaactc	ccctattcca	cttttattct	tacaatttag	cacttggtct	14820
cttctaacgt	attctaagac	ttactcattt	attacattgt	ttgccacccc	ctctagtaca	14880
taaaactccag	aggggcaggg	atttctgtct	atttattcat	ttctttatcc	ctaggacata	14940
gaacagggca	tagttcagag	tattcaatgt	tatcaatgaa	tgaactagca	gtagtaccag	15000
ttccagttag	gcacagaatt	aaatctaaat	agaattaaat	ctcatgggtc	gggttaacta	15060
tggaatagaaa	attagatata	attttaagaa	gcctagaaaag	aaaaaattaa	taatgtaaaa	15120
ataatattaa	tttgataata	ataacaaaaa	ctctgccagg	cactgtgggt	caaatctgca	15180
atcccagcta	ctcaggaggc	tgaggtggaa	ggatcacttg	agaccagagt	tcaagactca	15240
gcctaggcaa	cacggcaaga	aactgtctct	aaaaaaatta	aaacttaaat	ttttaaaaaa	15300
gaattctcaa	agcgtcacia	aaactggaga	ttaagggtaca	ggaagtgtga	agtaatatata	15360
ctatgctaata	gggttttttt	tttttttagaa	aggtataacc	aaaagatttc	tttctcaagt	15420
cgataaaactg	agaaagataa	gcataatctt	caattaacag	agggggagga	aaagccagat	15480
acaacaaaat	aagatatata	ttagtttcca	gtgaaaacaa	agagttaggag	ttattttgca	15540
tcacctcacc	tgtgacctcc	cccagcccaa	aaaacactac	tgataaacag	ggtagaaaag	15600
catcatctca	gataaagcag	gaaaaactgc	cacagtctca	aaccacaaac	tataagcaca	15660
cacctggcca	accctgccaa	gtctgggctc	agtaggagga	acgtgctgag	agctaggatg	15720
taccaactta	gacattctgt	gggatacaga	tgtccctgga	agggtcacac	catctcaag	15780
gcacctgtaa	tgccactga	ttacagccac	catatgtgag	agagaaactc	agggcactta	15840
gagagtataa	caagaacctt	atgtcatctg	agatgaggaa	tcctcagccc	tgcaaatata	15900

66/122

ccaactcttt	agaacaactg	gcaaaacata	aatatccaca	acttttgttt	cagtaattcc	15960
actcttagat	atcaatccaa	agtacatgag	acagcagata	cacacacaaa	atgggtattta	16020
ctgcagcatt	gtttataata	gcaaaaaaca	agaaataatc	catatgtctc	aataggatac	16080
tgggtacatg	agggtatgta	cccattcattc	aaccatcaaa	aagagtgtata	tggatgtcca	16140
cagatggaca	taaaaagctg	tgtgttacgt	gaaaacaaaac	tcaagcagca	gcaggatggg	16200
cttatgatag	tcagtatgag	ctaattttctg	gaaaaaaaaa	tctagtgtgt	gcacagaaaa	16260
catctgaaag	aacagaaaca	aaactatcag	cagaatattg	agatgtttta	ctaagttgta	16320
tatctatact	gcttgtaatt	tttaccctcaa	gcaagaatta	ctttttggaa	aaagaaaatt	16380
caggaaataa	agcattttctt	taaacttcat	gttttaacaa	atgggtgatgg	aataaaagag	16440
ttcttattca	tcataaacac	acacagcaca	catgcacgca	tgtgcgtgag	cacacccttt	16500
acttgataaa	taccatgttg	aatatttttag	tctttctctt	taggttctat	cccttcactc	16560
aaaatgcggt	tataaataaa	tgtacttttc	atgtgccttc	tgcttaaacc	cactttaata	16620
taactttaca	gtcccattat	catttatagtc	tcaaagctag	actcagcctg	aaactaccct	16680
ttcatttgga	acccttatta	aaatgccaca	tacagctcct	tcaataaaaa	acaaacccta	16740
ggacctgaca	ctaggcttcc	tttgttgcta	ctcataatgg	ccaagtctctg	tgcttataat	16800
acatcttctt	tcattttatt	gctacatata	caagggtttt	atatgttttt	cttattatat	16860
cttauttcaa	aacaccatca	cgctcttttc	cagatgaaaa	taaggaaaaag	aaattgagca	16920
actgactgac	ttaaagggtca	taaaactata	tagtagcaga	gtcagcaaaa	gaagaaacac	16980
adattctccc	agtagaggct	gaaaaccagt	accattcacc	tccagggtga	gctatatata	17040
guttacaaaag	tcaccttctc	taaatgttca	aactgaatcc	catacccata	ctttaccact	17100
acctcgtaag	aacagcctca	gatcttggtta	tagccttttt	tttagcatgc	tgaagccaat	17160
aanatgcttc	ccattcagca	agagaaacaa	gttctgaaac	actgaataat	ctgcccaggg	17220
cctatgaaca	tttccactgt	gagaaatggt	ctccactgtg	tggagaagat	ccttactctt	17280
ctccacacag	gcagaacatt	agaaaaattc	ttggattcta	tgatgcacag	cttaggagtc	17340
tgtttagcac	aattttaagtc	caaatagtta	ttaaatcctc	ctctgttcca	gaaacagtgc	17400
taaatactgt	gaatataaaa	attgaaaaga	tactctcctg	gctcccaaga	aagtcagcca	17460
gatagaggag	acacaggcac	acaaatcact	gtcacatgaa	gctctacctc	cctaacttca	17520
aacgagggcc	taagtcacca	agaatacagt	agcagttgtg	actacgagta	actactataa	17580
ttcaataact	tatcttccct	tagaaaactc	ttctcccttg	gaaatttatt	tgcatctcta	17640
aataccattc	cttactaaaa	ggaagcaggg	ctccttgggg	aaatagctga	ttctaggtgt	17700
ggactatgaa	atgaaaatgg	tgagtctggg	acatcccatg	ttgcccagaa	atcaagggaac	17760
tgcccagaag	ttaacagagt	catgttaaat	ggacctaaga	gtgaaccaga	aggagctcac	17820
tttgccccgc	ttcagaagaa	ttcaagaaaa	acatgcacagt	aatgaattat	aaaacatgaa	17880
ttaaaaataca	tatttggtact	aaaaagagaa	caaaaggatg	tggctttgga	taaagctctt	17940
cttcatggaa	gaataccagc	taataaatgt	aaaggaaatg	agagaattag	aaaaattatc	18000
attttgtaaa	ccttaatatata	ttcacctaga	catgctaaaa	ccactgagta	aaaggctgct	18060
tgggaagagg	atgctcacat	gatctcagag	tttcacacca	cagataattt	attagatata	18120
ggaagggaaga	tgtgatcaag	cttctgtgta	ccccagcca	ggccccacaa	cactatgtgc	18180
ctccttgtga	tgtgggagct	acacagcatc	gcccacacag	cttctcgcca	aaactgtttg	18240
aagctaatac	caagggaaga	actggacagc	ttctgaccat	gagacgctcc	accagacaac	18300
ttgcttggcc	cttccaaaga	aacttgcttg	gcctctccaa	agaaaactca	gtttcattta	18360
aaaacaaaac	taattattta	aaaacaaacg	aaaagcaagt	tgtggacttg	agctccaggg	18420
acagagcaga	catacttttc	cctgttcttc	ccagtaagtg	gtaataaaaa	ccctcaacac	18480
tagatataaa	acaaatataa	gaagggttctg	gaagggggaag	aggaggcaga	ctatccagggt	18540
gccttgaggc	ccacagaaca	acccagtgtat	gggttcaactg	ggctctcttt	ttgcttcatt	18600
atctcagact	tggagctgaa	gcagcaggca	acttcaaaac	accaaggggc	acagatgaa	18660
aagccccaag	aaaagcctgc	cctctctagc	caaaggacca	ggaaggagac	agtctaataa	18720
gatggaacac	atthagacag	taactgcccc	tttaccagca	ataactgagc	agggagccta	18780
gacttccagt	cttgtgagga	cgtaccaagg	tacccaacac	ccccaccaag	gctgagtaag	18840
gactgcgact	tttatccctg	catggcagta	gtaaggagcc	catccctcac	ccgccagcag	18900
tgtcagggga	acctggactt	ccactcccac	ccaggagtga	tgaggccctc	cctgctgggg	18960
tcattgtcaga	ggaggcctag	tggagattca	gtgacttaac	cttttccag	agataatgag	19020
gccacctttc	ctccctcttc	ccccatgggtg	acagtgaag	cactgtggca	agcagtaggc	19080
actcctaccc	ctcctagcca	gggaggtatc	agggaggcca	agtagggaac	cagaataccc	19140
acaaccaccc	agcagcaaca	ggggtccccc	acccccattgg	gtgtcaatgg	aagcagagcg	19200
gaaagcctgg	atattttacc	ccatctagaa	gtacaaagct	gatgtcccc	ttcttctact	19260
acaatgggtg	tcaaaacagg	tttaataaag	gtctagagtc	tgataacgta	atacccaaat	19320
cgttggaagt	ttcattgagg	atcattttata	ccaagagtca	ggaagatccc	aaactgaaag	19380
agagaaaaga	caattgacag	acactagcac	taagagagca	cagatattag	aactacctga	19440
aaggatgtta	aagcacatat	cataagcctc	aacaggctgg	gcgcgggtggc	tcacgcctgt	19500

67/122

aaccccgagca	cttttgggagg	ccgaggcgag	tggatcacaa	gatcaggaga	tcgagaccat	19560
cctgggctaac	acggtgaaac	cccgtctcta	ctaaaaatc	aaaaaaaaat	agcaaggcat	19620
gggtggggg	acctgtagtc	ccagctactc	gggagcctga	ggcaggagaa	tggcatgaac	19680
ctgggaagag	gagcagtgag	ccgagatcgc	accaccgcac	tccagcctgg	gcaacagagc	19740
aagacttcgt	ccccaaaaaa	aaaaaaaaaa	aaaaaaaaagc	ctcaacaaac	aactacaaac	19800
gtgcttgaaa	caaagtaaaa	aaaaatcttg	gcaaagaaat	aaaagatata	tattttggcc	19860
aggtgcagtg	gctcacagcc	tgtaatccct	gcactttggg	aggctgaggc	aggcggatca	19920
cctgaggtca	ggagtttgag	accagcctga	ccaacatgga	gaaaccccgt	ctctactaaa	19980
aatacaaaat	tagccagtc	tggtggcaca	tgcctgtaat	cctagctact	caggaggccg	20040
aggcaggaga	atcgcttgaa	ctcaggaggt	ggaggttgcg	gtgagccgag	atcccgccat	20100
tgacatttgc	actccagcct	gggcaacaag	agcaaaactc	catctcaaaa	aaatagatac	20160
atattttta	ggaaatttta	gaattgaaaa	atacagtaac	caaattgaat	ggaaagacaa	20220
catagaatgg	agggggcgaga	caaaataatc	agtgaacttc	aacagaaaa	aatagaaatt	20280
acccaatatg	aagaacagaa	agaaaaataga	ctggccaaaa	aataaagaag	aaaaaaggag	20340
agcagcagga	ggaatgatgg	aaaaagagaa	aggaagggaag	gaagggaagg	aggggaggga	20400
ggagtgaggg	agaaagtctc	aaagacctct	gagactaaaa	taaaagatct	aacacttgct	20460
atcaggggtcc	aggaaagaga	caaagatggc	acagctggaa	acgtattcaa	aaaataatag	20520
ctgaaaaactt	cccaaatttg	gcaagagaca	taaacctata	gattcgaaat	gctgaacccc	20580
aaataaaaa	cccaataaaa	tccacaccaa	atacatcat	agtcaaaact	ctgaaaagac	20640
gaaaagagaa	aacgtcttga	aagcagtgag	tgaacaaca	cttcatgtat	aagggaaaaa	20700
caattcaagt	aacagatttc	ttacagaaat	taaggaagcc	agaaggaaat	gacacaatgg	20760
ttttcaagt	ctgaaagaaa	agaagtgtca	acacaaaatt	ctagattcag	taaaaatatt	20820
cttcaagaat	caatgggaaa	tcaagacagt	ctcagataaa	gcaaaaataag	agaatatgtt	20880
gccagcagat	ctcccctaaa	ggaatggcaa	aaggaagatc	atgcaacaga	ccaaaaaatg	20940
atgaaagaag	gaatccagaa	acatcaagaa	gaaagaaata	acatagtaag	caaaaaatata	21000
tgtaattaca	ataaaatttc	tatctcctct	taagacttct	aaatttatatt	gatgggtgaa	21060
gcaaaaaatta	taaccctgtc	tgaagtgtct	ctactaaatg	tatgcagaga	attataaatg	21120
gggaaagtag	aggtttctat	acctcattga	agtggtaaaa	tgacaacact	gtgaaaagtt	21180
acatacacac	acacacgtaa	gtatatataa	atatatgtgt	gtatatgtgt	gtgtatatat	21240
atatatacat	ataatgtaat	acagcaacca	ctaacaacac	tatacaaaga	gataataacc	21300
aaaaacaatt	tagataaaat	gaaatggaat	tctaaaaaat	attcaaatat	tctacaggaa	21360
gacaagacaa	aaagagaaaa	aaagaggagg	acaaactaaa	tttttttaaaa	acataaataa	21420
aatggtagac	ttaagcccta	acttatcaat	aattacataa	atgtaaatga	tctaattata	21480
tcaattaaaa	gacagagata	gcagagttaa	tttaaaaaaca	tagctataag	aaacctgctt	21540
tgggctgagt	gcagtgactc	acacttgtaa	tcccagcact	tcgggaggcc	aaggcggtg	21600
gatcacctga	ggtcaggagt	tccagaccag	cctggacaac	atggtaatac	cccattctta	21660
ctaaaaaatc	aaaaaaatta	gccaggcatg	gtggcacacg	cctgtagtcc	caactactca	21720
ggaggctgag	acacaagaac	tgcttgaacc	cgggcgacag	aggtagcagt	gggccaagat	21780
tgcgccactc	cagcctgaac	gacagagtga	gactccacct	cagttgaaaa	acaaaaaaga	21840
aacctgcttt	aaatatacca	acatatgttg	gttgaaatta	aaagaatata	atatatcatg	21900
aaaacattaa	tcaaaagaaa	ggagtggcta	tattaataac	ataaaataga	cttcagagaa	21960
aagaaaaatt	caagagacag	gaataaaaagg	atcaagaaaa	gatcctgaaa	gaaaagcagg	22020
caaatacaat	attctgcttg	gagattcaac	accctctctt	aacaactgat	agaacaacta	22080
gacaaaaaaa	tcagcatgga	gttgagaaga	acttaacacc	actgaacaac	aggatctaata	22140
agacatttat	ggaacactct	acccaacaat	agcaaaaata	acattctttt	caagtattca	22200
ctgaacatat	ccttagaccc	taccctgggc	cataaaaata	agctcactag	tgattgccga	22260
aggcttggat	ggacagtggg	agagctgcac	gggggaggag	aagggtgacag	ttaaagagtg	22320
taggatttct	ttttgggata	atgaaaatgt	tccaaaattg	attgtggtga	tgttggcgca	22380
actctacaaa	tataaaaaag	gccattgaat	tgtacgtttt	aagtgggtga	aacatatggt	22440
atgtggatta	tatctaacgc	tttttaaaaa	cttaacacat	ttcaaagaat	agaagtcata	22500
cagagtgtgc	tctactggaa	tcaaaactaga	aagaggtaac	tggaggataa	cgagaaaagc	22560
ctccaaatac	ttgaaaactg	gacagcacat	ttctaaaatc	atccgtgggt	caaagatatt	22620
catttctgat	attcattttt	attgtttta	gtatttttaa	aaatttctta	agggaaataa	22680
actgactaaa	aatgaatatg	gctgggtgag	gtgggtcacg	cctgtgatcc	cagcactttg	22740
ggaggccgag	gctgggtgag	cacaagatca	ggagttcgag	accagcctgg	ccaagatggt	22800
gaacccccgt	ctcaactaaa	aaactacaaa	aagtagccaa	gcgcagtggc	gggagcctgt	22860
ggtcccagct	acttgggagg	ctgaggtagg	agaatcgctt	gaacacaggc	agcagaggtt	22920
gcagtgagcc	aagattgtgc	cactgcacgc	cagcctgggc	gacagagact	gcctcaaaaa	22980
aaaaaaaaaa	aaaaagaata	tcaaaatttg	tgggacatag	ttaaagcaat	gctgagaggg	23040
aaatttataa	cactaaatgt	ttacattaga	aaagagaaaa	agtttcaaat	caatagtctc	23100

68/122

cactcccatc	tcaagaacac	agaagatgaa	gagcaaaata	aacccaaagc	aagcaaaaaga	23160
aagaaaatat	aaaaataaat	cagtaaaatt	gaaaacagaa	acacaataaa	gaaaatcagt	23220
gaaacaaagt	actgattctt	cgaaagatta	ataaaattga	caaacctcta	gcaaggctaa	23280
caaacaaaaa	agaaagaaga	cacggattac	cagttattag	aatgaaagca	taattagaaa	23340
caactctaca	cattataaat	ttgacaatgt	agatgaaatg	gactaattac	tgaaaaaaca	23400
caaattacca	caactcacc	aatatgaaat	agataattgg	gatagcctga	taactactga	23460
gaaaattgaa	tttghtaattt	taacactctt	aaaacagaaa	cattaaactt	aatattttat	23520
aaatattaga	taaggtaatt	atacccttcc	ttacacaaata	aaaacgacaa	attattttgc	23580
agctaaagag	atgtatgtac	tgtgaaaaat	atcttcagaa	aaatagaact	ttgtttgaag	23640
aataaggatt	taaaaaatgt	ttttaactct	caagaagcaa	atatctgggc	ccagatgggt	23700
tcactgaaga	attctaccac	atgtttaatg	aagaattacc	accaactcta	catagcatct	23760
ttgagaaaac	tgaagagaag	ggaacatctc	ccagttcatt	ttatgaagtg	gggtgttactc	23820
tgatactaga	actgtataag	gacagctact	cttgacacac	tgcttatggg	tagctctgct	23880
ctgcaggaa	agtgcagaaa	aaaaaaaaaa	gaagcactgg	acaagggcag	tataaaaaaa	23940
gaaaactggg	ccaggtgcag	tggtctcacac	ctgtaattctc	agcactttgg	gaggctgacg	24000
ctgggtggatc	acctgaggtc	aggagtttga	gactagcctg	gccaacatgg	taaaaccttg	24060
tctctactaa	aatacaaaaa	ttagccaggc	agggtgggtg	ggaaaaataaa	aaggaaaaaa	24120
aaacaaaaat	aaactgcaga	ccaatatcct	tcattgagtat	agacacaaaa	ctccttaaac	24180
tccttaacaa	aattattagca	agtagaagca	ataataaaaa	ataattatac	accatgcatc	24240
agtgggactt	attccagaaa	cgcaagtctg	gttcaacatt	tgaaaacaag	gtaaccact	24300
atatgaucgt	actaaagagg	aaaactacat	aatcacatca	atcaatgcag	aaaaaagcat	24360
ttgccaaaat	ccaatatcca	ttcatgatac	tctaataaga	aaaataagaa	taaaggggaa	24420
attccttgac	ttgataaaagc	ttacaaaaaga	ctacaaaagc	ttacagctaa	cctatactta	24480
atgggtgaaaa	actaaatgct	ttcccctacg	atcaggaaca	aagcaaggat	gttcactctc	24540
attgctctta	tttaacatag	ccctgaagtt	ctaacttggtg	caaaacgata	agaaagggaa	24600
atgaaagacc	tgacagattgg	caaagaagaa	ataaaaactgt	tcctgtttgc	agatgacatg	24660
attgtctcat	agaaaatgta	aagcaactag	gggtaggggg	gcagtggaga	cacgctgggtc	24720
aaaggatacc	aaatttcagt	taggaggagt	aagttcaaga	tacctattgc	acaacatggt	24780
aactatactt	aatatatgtt	attcttgaaa	ataactaaaag	agtgggtgtt	aagcgttctc	24840
accacaaaaa	tgataactat	gtgaagtaat	gcatacgtta	attagcacia	cgtatattac	24900
tccaaaaaat	catgtttgtac	atgataaata	cacacaattt	tatctgtcag	tttaaaaaa	24960
catgattttg	gccaggcaca	gtggctcata	cctgtaattcc	cagcatttta	ggaggctgag	25020
gcgagcagaa	aacttgaggt	cgggagtttg	agaccagaat	ggtcaacata	gtgaaatccc	25080
gtctccacta	ataatacaaaa	aattagcagg	atgtgggtggc	gtgcacctgt	agaccagct	25140
acttgggagg	ctgaggcacg	agaattgctt	gaacaaggga	ggcagagggt	gcagtgaagt	25200
gggtggccact	gcattccagc	ctgggtgacag	agtggagactc	catctcaaaa	aaaataaaaat	25260
aaagcatgac	ttttctttaa	tgcaaaagcag	ccaagcgag	tggtcatgc	ctgtaatccc	25320
accactttgg	gaggccgagg	caggcagatc	acaaggctcag	gagtttgaga	ccagcctgac	25380
caacatgggtg	aaaccccatc	tctactaaaa	aatatataaa	ttagccaggc	atgtgtagtc	25440
tcagctactc	aggaggctga	ggcaggagaa	tcacttgaac	ccggaggcag	aggttgcagt	25500
tttgagccac	cgcactccag	cctgggtgag	agaacgagac	tccgtctcaa	aaaaaaaaaag	25560
caaaaataacc	taatttttaa	aacactaaaa	ctactaagtg	aattcagtaa	gtctttagga	25620
ttcaggatat	atgatgaaca	tacaaaaatc	aattgagctg	gacaaaggag	gattgtttta	25680
ggtcagtagt	ttgaggctgt	aatgcacaat	gattgtgcct	gtgaatagct	gctgtgctcc	25740
agcctgagca	gcataatgag	accacatctc	tatttaaaaa	aaaaaaaaatt	gtatctctat	25800
gtactagcaa	taagcacatg	ggtactaaaa	ttaaaaacat	aataaaact	gtttttaatt	25860
gcctgaaaaa	aatgaaatac	ttacatataa	atctaacaaa	atgtgcagga	cttgtgtgct	25920
gaaaactaca	aaacgctgat	aaaagaaatc	aaagaagact	taaatagcgt	gaaatatacc	25980
atgcttatag	gttggaatac	ttaatatagt	aaagatgcca	attttatcca	aattattaca	26040
caggataaca	ttattactac	caaaaatccc	gaaaataggt	acatagatat	agacaagatc	26100
atacaaaaaat	gtatacggaa	atatgcaaag	gaactagagt	agctaaaaaca	aatttgaaaa	26160
agaaaaataa	agtgggaaga	atcagctctat	ccagtttcaa	gacttacata	gctacagtaa	26220
tcaagactgt	gatattgaca	gagggacagc	tatagatcaa	tgcaaccaaa	tagagaacta	26280
agaaagaagc	acacacaaat	atgccccaat	gattttctgac	aaagggtgta	aaacacttca	26340
acgggggaag	atatgtctct	cattaaaggg	tgttagagtca	ttgcacatct	ataggcaaaa	26400
agatgaacct	gaacctcaca	ccctacagaa	aaattaaactc	aaaatgactc	aaggactaaa	26460
cataagatat	acatctataa	aacatttaga	aaaaggccac	gcacgggtggc	tcacgctcgt	26520
aatcccagca	ctttgggagg	ccaaggcagg	tggatcacct	aagggtcagga	gtttgagacc	26580
agccggatca	acatggagaa	gccccatctc	tactaaaaat	acaaaattag	ctggacgtgg	26640
tggcacatgc	ctgtaatccc	agctacttgg	gaggctgagg	catgagaatc	gcttgaaacc	26700

69/122

gggggggcaga	ggttgcggtg	agccaagatc	acaccattgc	actccagcct	gggcaacaag	26760
agcaaaactc	caactcaaaa	aaaaaaaaaa	aaaggaaaaa	tagaaaatct	ttgggatgta	26820
aggcgaggta	aagaattctt	acacttgatg	ccaaactaag	atctataagg	ccagtcgtgg	26880
tggctcatgc	ctgtaattcc	agcactttgg	tcaactagat	gaaaggtata	tgggaattca	26940
ctgtattatt	ctttcaacct	ttctgtaggt	ttgacatttt	tttagtaaaa	aattggggga	27000
aagacctgac	gcagtggctc	acacctgtaa	tcccagcact	ttgggaggcc	ggggcgagtg	27060
gatcacacgg	tcaggagttc	gagaccagcc	tggccaacat	ggtgaaaccc	cgtctctacc	27120
aaaaatataa	aaaattagcc	gggtgctcatg	gtgcatgcct	gtaattcccag	ctactgagga	27180
ggctgaggga	ggagaatcac	ttgaacctgg	gagggtggaag	ttgcagtggag	ccgagattgt	27240
gccactgcac	tccagccttg	ggtgacagag	cgagactccg	tctcaaaaga	aaaaaaaaaa	27300
aaagaatatc	aaacgcttac	tttagaaact	atttaaaggga	gccagaattt	aattgtatta	27360
gtatttagag	caatttttat	gctccatggc	attgttaaat	agagcaacca	gctaacaatt	27420
agtggagttc	aacagtgttt	aaatttgcta	actggtttagg	aagagagccc	tatcaatatc	27480
actgtcattt	gaggctgaca	ataagcacac	ccaaagctgt	acctccttga	ggagcaacat	27540
aagggggtta	accctgttag	ggtgttaatg	gtttggatat	ggtttggttg	gccccaccga	27600
gtctcatggt	gaaatttggg	ccccagctact	ggaggtgggg	ccttattgga	aggtgtctga	27660
gtcatggggg	tggcatactc	ctcctgaatg	gtttgggtgcc	attcttgcag	gaatgagtga	27720
gttcttactc	ttagttccca	caacaactgg	ttattaaaaa	cagcctggca	ctttcccca	27780
tctctcgctt	cctctctcac	catgtgatct	cactgggttc	ccttcccttt	atgcaatgag	27840
tgggaagcag	ctgaagccct	cgccagaagc	agatagtgat	gccatgcttc	ttgtacagcc	27900
tcaaaaacca	tgagcccaat	aaaccttttt	tctttataaa	tattccagcc	tcaggatttc	27960
ctttatagca	agacaaatga	accaagacag	ggggaaatca	acttcattaa	aataatctat	28020
gcagtcacta	aacaaataag	aacaagaggc	tccagaagtg	ggaagccaat	accagaggtt	28080
cctacaatac	agtatctgaa	aagtccagtt	tccaacccaa	aaatatatat	atacaggccg	28140
gacatggtag	cttatgtctg	taatcccagc	actttgggat	gctgaggctg	gcagatcacc	28200
ctaggtcagg	agttcgagac	cagcctggcc	aatatggcca	aaccccgctc	ctactaaaaa	28260
tacaaaaaatt	agccaggcat	ggtggtggat	gcctgtaatc	ccagctactc	gggaggctga	28320
ggcagggaat	cacttgaacc	caggaggcag	aggttgcagt	gagccgagat	cacgccactg	28380
aactccagcc	tgggcaacaa	agtgagactc	cacctcaaaa	aaaaaaaaaa	tatacatata	28440
tatatgtgtg	tgtgtgtgtg	tgcgcgcgtg	tgtgtatata	cacatacaca	tatacatata	28500
tatacagaca	cacatatata	tatgaagcat	gaaaagaaac	aaggaaagtat	gaaccatact	28560
ttctgtggtt	atgataggat	gggggtatcac	gggggaagta	gacaagggaa	actgcaagtg	28620
agagcaaaaa	gttatcagat	ttaacagaaaa	aaagactttg	agtaaccatt	ataaatattg	28680
ccacagaatt	aaagaaaagc	gtgattaaaa	aaggaaaagg	aagtatcata	acaattattac	28740
tccaaataga	gaatatcaat	aaaggcatag	aaattataaa	atataataca	atggaaattc	28800
cggagttgaa	aggtagaata	actaaaattt	aaaattcact	agagaagggt	caacactata	28860
tttgaactgg	cagaagaaaa	atttagtgag	acaaatatac	tccaattagac	attattcaaa	28920
tj.aaaaata	aaagaaaaaa	gaatgaagaa	aaataaacag	aatctcagca	aaatgtggca	28980
cuccattaat	cacattaaca	tatgcatact	gagagtaccg	gaagcagatg	agaaagagga	29040
agaaaaaata	ttcaaattgat	ggccagtaac	ttcctagatt	tttgttttta	agcaataacc	29100
tatacaatca	agaaaactcaa	tgaattccaa	gtaggataaa	tacaaaaaga	accacaacaa	29160
gatacaccat	ggttaaaaaatg	ctgtaagtca	aaaaacagaa	aaatattgaa	agcagctaga	29220
ggaaaaactta	taagagaacc	tcacttcaaa	aagaacatca	cttataaaaa	aaccacaata	29280
atagaaacag	ttgacctctc	atcagaaaca	atgaatgata	acatatattga	agtgtctaaa	29340
gaaaaaaaat	aaagatttctg	atatacgaca	aaagtgtctt	tcaaaaatat	acatccaaaa	29400
ggattgaaac	cagggtctct	aagagttatt	tgtacatcca	tgttcatagc	agcattattc	29460
acaatagcca	aaaggtagaa	gcaacccaag	ggtccatcga	caaatataata	aaatgtggta	29520
tatgtataca	caatggaatt	tattcagtat	taaaaaggaa	tgaatttctg	acacatgcta	29580
caacatggct	aaacctttgag	aacactatgc	taagtgaatt	aagccagcca	caaaaggaca	29640
aataccatat	tacttcaatt	gtatgaaata	cctagggtag	tcaaattcag	agatagaaa	29700
taaaacagtg	gttgccaagg	gctgaggagg	ggagtaacgt	ggagtatttg	ttgaattgggt	29760
acagaatttc	agttttgcaa	gataaaaaaga	gttctggaga	cagatgggtg	tgggggtgg	29820
acaacaatac	aaatatactt	tatactactg	aacagtatac	ttaaaaatga	ttaacatggt	29880
gaaaccccg	ctctactaaa	aatataaaaa	aattagctgg	gtgtgggtgg	gggcacctgt	29940
aatcccgact	acttgggagg	ctgaggcagc	agaattgctt	gaaaccagaa	ggcggagggt	30000
gcagtgagct	gagattgcgc	caccgcactc	tagcttgggc	aataagagca	aaactccgtc	30060
tcaaaaaata	aaaaataaaa	aaaattttaa	aatgattaa	caggaggcca	ggcacgggtg	30120
ctcacacct	taagtgccag	actttgggag	gccgaggcag	gcgatcactt	gagaccagga	30180

70/122

tgcttgaacc	caggaggcag	agattgcagt	gagtcgagat	cgcgccactg	aattccagcc	30360
tgggcgcag	agcaagattc	tgtctcgaaa	aaacaaaaac	aaaaacaaaa	agcaaaacca	30420
aaaaataatt	aagcaggaaa	cgagattgct	gctgaggagg	agaaagatgt	gcaggaccaa	30480
ggctcatgag	agcacaaaac	ttttcaaaaa	atgtttaatg	attaaaaatg	taaattttat	30540
atgtatctta	ccacaaaaaa	aagggctggg	gggcaggaaa	tgaagggtgaa	ataaagacat	30600
cccagagaaa	caaaagtaga	gaatttggtg	ccttagaaga	aacaccacag	gaagttcttc	30660
aggctgaaaa	caagtgaccc	cagagggtaa	tctgaattct	cacagaaaat	tgaagcatag	30720
cagtaaaggt	tattctgtaa	ctatgacact	aacaatgcat	atTTTTTcct	ttcttctctg	30780
aaatgattta	aaaagcaatt	gcataaaaata	ttatatataa	agcctattgt	tgaacctata	30840
acatatatag	aatataactt	gtaatatatt	tgcaaaatac	tgcaaaaaag	agagttggaa	30900
caaagctggt	actaggctaa	agaaattact	acagatagta	aagtaatata	acagggaact	30960
taaaaataaa	atttttaaaa	atttaaaaaa	aataattaca	acaataatat	ggttgggttt	31020
gtaatattaa	tagacataat	acaaaaatac	cacaaaaagg	gaagaagaca	atagaactac	31080
ataggaataa	cattttggta	tctaactaga	attaaattat	aaatatgaag	tatatctctg	31140
taagttaaga	cacacatggt	aaaccctaga	tactaaaaag	taactcacat	aaatacagta	31200
aaaaaataaa	taaaataatt	aaaatgtttg	tattagtttc	ctcagggtac	agtaacaaac	31260
taccacaaat	tgagtggctt	aacacaactt	aatgtatttt	tctcccagtt	ctggaggcta	31320
aacacctgca	atcaaggtga	gtacagggcc	atgctccctg	tgaaggctct	aggaaagaat	31380
cctcccttgt	ctcttcagc	ttccagtggt	tctcagtaac	cctaagtgt	ccttggcttg	31440
tagctatatt	attcctagca	accagaaaga	agaaaataat	aaagattatg	gcaaaaaata	31500
atgaaatcaa	aaggagaaaa	atggaaaaaa	ataaataaaa	ccaaaagcta	gttctttgaa	31560
aagatcaacc	aagttaacaa	accttttaac	tagactgaca	aaaaggaggt	aagactcaaa	31620
ttactagaat	cagaaataaa	agagggggaca	ttactaatga	gggattagaa	agaataacta	31680
cgaacaaatg	tgtgccaaca	aattagaaaa	ccttagatgaa	atggacaggt	tcctaggaca	31740
acatcaacta	ccaaaattta	ctcaagaaga	aagagacaat	ttgaatgagc	tataacaagg	31800
gaagagactg	aattgacaac	caagaaacta	tccacaaaga	aaatcccagg	cccagaagat	31860
ttcactgtga	aattctttca	aacttataaa	tataaattaa	catcagttct	tcacaaactc	31920
ctccaaaaaa	aagaacagat	ctctattttac	aggcgatagc	atcttttagaa	aatcctaagg	31980
gaactactaa	gacactatga	taactgataa	acaagttcag	caaggctgca	ggatagaaaa	32040
ccaatataca	aaaatctatt	atattttctat	acacttgtag	tgaacaaccc	aaaaatgaga	32100
ttaagaaaaa	aattcaattt	acaataacat	caaaaagaat	aaaaacactc	aaaaataaat	32160
ttattcaagt	aagtgcaaaa	cttatactct	agaagctaca	aaacactggt	aaaagaaatt	32220
aaaggtttac	ataaatgaaa	aactatccca	tgttcatgga	tcaaaagact	tattactggc	32280
aatgctctcc	aaattgatct	ataaattcaa	caaaatcctt	atcaaaatcc	cagatgaggc	32340
tgggggtggc	ggttcatgcc	tgtaatccca	gcactttggg	aggctgaggc	acgcagatta	32400
cctgaggtcg	ggagctcgag	atcagcctga	ccaacatgga	gaaaccctat	ctcttctaaa	32460
aatacaaaat	tagtcaggcg	tgggtggcaca	tgccataat	cccagctact	cgggaagctg	32520
aggcaggaga	atcgcttga	cccaggaggc	agaggttgca	gtgagccaag	atcggtgccat	32580
tgcactccag	cctggggcaac	aagagcaaaa	ttccacttca	aaaaaaaaaa	aaaaaaaaatc	32640
ccagatgact	tcaactgttga	aattgaaaag	attattctaa	aattcacatg	gaattgcaag	32700
accttgagaa	tagccaaaaac	aaacttgaaa	aacacgaaca	aatatatagga	tgactcactt	32760
gccaattgca	aatgttacga	cacagcaaca	gtaatcaaga	ctgtgtggta	ctggcaaaag	32820
acacatacat	acatacatat	caatggaata	taattgagag	tacagaaaca	agcctaaaca	32880
tctatggtaa	gtgctttttct	atTTTTTtct	TTTTTTTTtt	ctTTTTTtga	gagatagaat	32940
ctcaccatgt	tgcccaggct	ggtcttcaac	ttctgggctc	aagcaatcct	cccactgtgg	33000
cctcccaaag	tgctgggata	actggcatga	gccaccacat	ccagcccaga	tgattttcaa	33060
aaaagtcaac	aagaccattc	ttttcaacaa	ataggtctgg	gatgatcaga	tagtcacatg	33120
aaaaaaaaaa	tgaagttgga	ccctccatca	cactaaagtg	ctgcgattat	aggcatcagc	33180
caccacatcc	agcccaaagt	atTTTcaaaa	aggtcaacaa	gaccattctt	ttcaacaaat	33240
aggtctggga	taatcagata	gtcacatgaa	aaaaaaaaatg	aagttggacc	ctccatcaca	33300
ccatatgcaa	aaaatlaattc	aaaaatgaat	tgatgactta	aacgtaagag	ttacgactgt	33360
aaaactctta	taagggaaca	tacgggtaaa	tcttaaagac	gttaggtttg	acaaagaatt	33420
cttagacatg	acaccaaag	catgaccaac	taaggtaaaa	tagggtaaat	tgtacctacc	33480
aaaatgaaaa	acctttgtgc	tggaaaggac	accatcaaga	aatggaaagc	caaaatagcc	33540
aaggcaatat	taagcaaaaa	gaacaaagct	ggaggcatca	tactacctga	cttcaaagca	33600
acagtaacca	aaacagcatg	gtactagtag	aaaaacagac	acatagacca	atggaaacaga	33660
ataaagaacc	caaaaaataa	tccacatatt	tatagtcaac	tgatttttga	caatgacacc	33720
ccttcaataa	atgatactag	gaaaactgga	tatcgatatg	cagaagaata	aaactagacc	33780
cctatctctc	accatataga	aaaatcaact	cagactgaat	taaagacttg	aatgtaagac	33840
ccaaaactat	aaaactactg	gtagaaaaca	taaggaaaaa	cgcttcagga	cattggtcca	33900

71/122

ggcaagatc	ttatggctaa	aacctcaaaa	acacaggcaa	caaaaacaaa	aatggaaaaa	33960
tagcacttta	ttaaactaaa	aagctcctgc	acagcaaagg	aaacaacaga	atgaaaaagac	34020
aacctgtaga	atggggagaaa	atatttgcaa	actatccatc	catcaaggga	ctagtatcca	34080
gaacacacaa	gtgactaaaa	caactcaaca	gcaaaaaagc	aaataatctg	gtttttatat	34140
gggcaaaaaga	tctgaataaa	cattctcaaa	ggaagacata	caaatgtcac	tatcattctg	34200
ccagtaccac	actgtcttga	ttacttgcta	gtgtataaat	ttttaaatgt	ggaagtgtga	34260
gtcatcctac	actttgttct	tgtttttcaa	gtttgttttg	gctattctgg	gagccttgca	34320
agtataaaat	agccaacaag	tatgaaaaaa	tgctcaccat	cactaatcat	cagagaaata	34380
aaaatcaaga	ccactatgag	atatcctctc	actccagtta	gaatggctac	tatcaaaaag	34440
acaaaatata	atggatgctg	gcaaagattt	ggagaaaggg	gaactcctat	acactgtggg	34500
tagggatgca	aattggtaat	ggccattatg	gaaaataata	ctgaggtttt	tcaaaaaact	34560
gaaaaataga	ctaccatatg	atccagcaac	cctactactg	ggtatttatc	caaaggaaaag	34620
aagtcagtat	actgaagaaa	tatatgcact	ctcatgttaa	ttgcaacact	gttcacaaca	34680
gccaagacag	ggaataaatc	taaatgtgca	tcaacagatg	aatggataaa	gaaaaatgtg	34740
catatacact	caatagaata	ctattcagcg	atataagaag	aatgaaatcc	tgtcattcca	34800
gcaacattgga	tgaacctgga	ggacattata	tttaatgaaa	taagtaaagc	acaaaaagat	34860
aaacagtaca	tgttctcact	cagacatggg	tgctaaaaag	aaaatggggg	cacagaatta	34920
gaaggggagg	cttgggaaaa	gttaatggat	aaaaattttac	agctatgtaa	gaagaataag	34980
ttttagtgtt	ctatagaact	gtagggcgag	tatagttacc	aataacttat	tgtacatggt	35040
caaaaagcta	gaagagattt	tggtgttcc	cagcacaagg	gaatgataaa	tgtttgtgat	35100
gatggatate	ctaattaccc	tgattcaatc	attacacatt	gcatacatgt	atcaaattat	35160
cactctgtac	ctcataaata	tgtataatta	ttacgtcaac	aaaaaaagga	aaaaaaagaa	35220
aattaagaca	acccataata	tgggaagaaat	aaaatatctg	caaattatat	atatctgata	35280
aataattta	atattataata	tataaagaac	tcctacaact	caagaacaac	aacaaaaaca	35340
cccaattcaa	aaatgggtaa	aagccttgaa	tatacactta	tctaaagact	atatacaatt	35400
ggccaataaa	gacacgaaaa	gatgctcaac	atcactagtc	atcagggaag	tataaatcaa	35460
aaccacaatg	tagaatgtag	acaccacttc	atatgcacta	ggatggctag	aataaaaaag	35520
taataacaaa	tggtggtaag	gatgtgaaaa	aatcagaaac	ctcattcgct	gctgttgga	35580
atgtaaatg	atgcagccac	tttggaaaaac	agtctggcag	ctcctcaaat	tattaaatac	35640
agagttaccg	tatgaccag	gaatattcct	cctgggtcta	taaccaaaaa	aatgaaaaca	35700
tatatccaca	taaaaacttg	tacatgggca	tttatagcaa	cattattcat	aacagcaaa	35760
gtggtaaaga	cccatatgcc	catcatctga	tgaacaggta	aataacatgc	ggtattatcc	35820
atacactaga	atattatctg	cccatacaag	gagtgacatc	cagctacatg	ctacaaggat	35880
gaatctcgga	aaccttatgc	taagtgaag	aagccagtc	caaatgacca	cagattatga	35940
ttccatgcat	cggaaatgac	cagaatagg	aaatctatag	agacagaaag	tagattagtg	36000
gttgggtggg	gctgggagga	caggtagtac	actactttcc	cagaactact	ggaacaaagt	36060
accacaact	ggggagctta	aacatagaaa	ttgatttctc	cacagtcttg	gagactagga	36120
ctctgagatc	agggtgtcag	cagagctggg	tctttctgag	ggcctgagg	caaggctctg	36180
tcccaggcct	ctctccttgg	ctggcagggt	gccatcttct	ccctgcgtct	tcacatcatc	36240
ttttctctgt	gtgtgcccac	gtccaaattt	tgattggctc	attctgggtc	atggccaatt	36300
gctatgcaca	aagtgaagtc	tacttccaaa	agaagggaag	agggaaacact	gactaggcta	36360
aacttatagt	catttttaatg	tccgcttttc	ctatgagatt	gtgaacacac	agaagtagg	36420
tttttatcta	cattgtgcaa	agtttaataa	gaaaaataga	attcaagaga	agcagttcaa	36480
tagcaggaat	ttaatatggg	aactaattac	aaggtttagg	gcaggactaa	aaagccagtt	36540
gggatgggtga	gccaaaccag	agattagcaa	cagtgggacc	ccatctacct	accacccatg	36600
aagctggaag	gataaaggag	gggctattat	cagagtcac	aagccagtg	cagagtcctt	36660
ggctggagct	gggaccacc	tagagacact	gtgcaaaagca	gaaaacaagg	gggaaaaaac	36720
ctgacttctc	ccttccctcc	acctttcaat	ctcccactag	tgcttctctac	tagccatact	36780
tggccagaga	cagtgaacag	gaacactgca	aatgaagtt	tgtaggaatc	atctccctct	36840
gagacagaga	aatatggaag	ggtagaaaat	gaatcagagg	ataaagagaa	aaaaccctga	36900
gtactactct	atttctcttt	gtatctccag	tgccataatc	gtctctcaaa	aaaggaaagc	36960
aattgagaga	aactgaaaac	tccaattgaa	atgaaagaat	ggagaattac	tggcatagaa	37020
gagaagagaa	aaatttatct	cgcatagagt	aaacaagaat	ggattcacia	aggacgtgat	37080
gaatgaaaag	ctataatcag	caaagatttg	ccagagaaat	taaaaagtgg	taaactcagc	37140
cacgctgtac	aacctgaagg	cacaatgcat	gaaaacgttt	caagaaatga	caagatttga	37200
agtcacattc	taagtgaatt	tccagaatct	ctcaagacga	ttatatagct	acccattttt	37260
attaaataaa	atggaaactt	actaaacttt	ccccttgat	taaactaaca	tatgtcctaa	37320
tagcaaacga	ttctggaatt	cctagagtaa	aatatatatt	gtcaaagtgt	attgtctctt	37380
taatattctg	ctgacctcct	tttgctattt	aggatatttg	tatacacatc	acacgtaaat	37440
ttggtctata	gtttacatct	acgggcttat	actgttcttt	ttttcatttt	tttaaaattt	37500

72/122

ccaacccccca	gstatccatat	actgctctct	atcaggggta	ttttaacttt	gtaaaatcag	37560
ctgagatgct	ttccatgttt	ttttttttta	ttttctgcca	catttgaata	gcataggagt	37620
taccaccatc	aaccttggtat	tatttaagca	ttcacgattc	cacgtgtgga	ttttttattc	37680
agagtctttc	ttgtcattcc	tgctatcagc	acagaaccca	atctcagctt	tccagctata	37740
ctctcaccac	atggaatttg	cagatgaagt	tcaaaaggac	ctttgcatta	tcctgcctcg	37800
ccctcttccc	ccttcattta	gacatcacct	tcttctagaa	cgtcttacct	gacatgccct	37860
gctcccaacc	cctgctgccc	aattgtgtgc	tctcccgtgt	cctggcctgc	catcctcttt	37920
agtaattgcc	tgctccctca	tctgtctccc	cacccagaca	ttaagctgaa	tagactggat	37980
ttgtgtcttg	tccatcacta	taatctcagc	acctagtacc	tagtaggtac	ttaccatgta	38040
ttcattagca	aaatgttatg	tataaccttg	caccttaaaa	acaagagaag	gaagacaaaa	38100
ttaagtctta	agactatggt	ttagaacatg	gacagaaaac	tacagtctgc	agcccaaatc	38160
cagaccaa	gaagagacca	tgttcattta	catacaacct	atagcagctt	tcacactaca	38220
ggagcagagc	taagtagtct	caagggaaca	cacggccctg	caaagcctaa	aatattttact	38280
ctatagctct	tcacagaaaa	agttttcaga	tccctcgttt	agaactcttg	ttcatatgca	38340
atttcactaa	accatagttt	tttgggtttg	tttggttttt	tttggcaaaa	aggaatgagc	38400
cgatccagaa	aagggtgaaa	agaatgaatc	attactgtctg	aaagaatgtg	cacacagctc	38460
gtcagtatcc	tgctgccatg	ctgacaccca	tccaatagtg	tcatgagatg	cagcagctac	38520
tactgtgttc	tcaatgccga	gtccacccac	tccataacca	tgtccaagca	atcttgggaa	38580
catcatcacc	atgcttgttt	atccttaagg	tattgcctca	catacagcag	tggtcggtca	38640
taaagtcaaa	tgacactagt	ggccaggagg	tcaagagaat	gagtggaggac	aggtgggtag	38700
gcagcccagg	ccctagcaac	agcaggagct	cacccctcag	tactcttagc	caggactgaa	38760
atacttttca	ccctttcaag	agagactagg	aatctggatt	tttatgtgaa	atatcttgat	38820
tactaaatgt	tgtcaacaga	catgtcaaaa	ggtaaaaacta	agtaagttca	tggggcagat	38880
tgactattca	ggttatgaa	ttaaggattc	ttatccaaca	cagataccaa	ccaaaaagct	38940
gacgtataac	atattagtag	aaactatgtg	cactgtcgaa	acatcaacaa	ggggctaatg	39000
tctaaaatag	tctatattgg	attccagttg	aaacatgggg	aaaggacatg	aacaggcaac	39060
ttatgtcaat	ggaaactcaa	aaagataaca	agcatatata	aaagcattct	caaattcagt	39120
agtaaacaga	cagatgcaaa	taaaaagagg	gaaactgctg	ccgggcacag	tggtctcacac	39180
ctgtaatccc	agcactttgg	gaggccgagg	cgggcggtac	atgaagtcag	gagatcgaga	39240
ccatcctggc	taacatggtg	aaaccccgctc	tctactgaaa	acacaaaaaa	ttagccaggc	39300
gtagtgggtg	gcaccagtag	tcccagctac	tcaggaggtt	gaggcaggag	aatggcatga	39360
acccaggagg	cggagattgc	agtgaaccga	gaccatgccca	ctgcactcca	gcctggggcga	39420
ctgagtga	ctccatctca	aaaaatatata	taataattat	aattataata	ataataata	39480
gtaataaat	aaaaagagag	agactgctaa	agtctagaaa	gttgaatgat	gccaagcgca	39540
tgcaaagatc	agggccttgg	gatggccggg	tgcatgtggt	cacgcctgta	atcccaccac	39600
tttggggaggc	caaggcgggc	ggatcatgag	gtcaagagat	caagaccatc	ctggccgaca	39660
cagtgaacc	cggtctctac	taaaaagtaca	aaaaaatata	tatatatata	tatatatta	39720
tattatata	atatatatca	gagccttggg	aatccttgtg	tgctgctggg	gaaggtagtg	39780
gtgcagccac	ccttgacagc	aatctggcag	tacttgggtta	tattaagtat	aggcacacac	39840
cacgaccagg	cagtccact	cctgggtcta	aatcccaaag	aattctcaca	caagtccata	39900
aggagacatg	tacgaggctc	attcagcatt	actgggagtg	ggaatcaacc	tggtgttcca	39960
tctacaggag	acgagatgga	caaaatgtgg	tgatatttaa	gaccagaatc	accaagtaac	40020
agagatgggt	ggtgagtga	aatcctaaga	tacagaataa	aggctagaac	atgatgccat	40080
tcatgtaaat	taaaaataga	tgacacacaa	gcagtatacg	cgtgaccctt	gaatagcaca	40140
ggtttgaact	gcctgtgtcc	acttacatgt	ggattttctt	ccacttctgc	tacccccaa	40200
acagcaagac	caacccctct	tcttccctct	ccccctcagc	ctactcaaca	tgaagatgac	40260
aaggatgaag	actttttatg	taatccaatt	ccaagggaact	aatgaaaagt	atattttctc	40320
ttccttatga	ttttctttat	ctctagctta	cattatttcta	agaatatggg	acataatata	40380
catcacacgc	aaaataaatg	ttaattgact	gtttatatta	tgggtaaggc	ttccactcaa	40440
cagtaggctg	tcagtagtta	agttttggga	gtcaaaagtt	atacacagat	tttcaactgt	40500
gcaggcaatc	agttccccctg	acccccctcat	tgctcacggg	tcaactgtat	atacacaaaa	40560
gtattatatg	aacctcatta	gaatagctgt	ctataggagg	aagagaatga	gagtgaggata	40620
aaacgggaatg	aacaaataaa	ccaacaaatg	cattaacaag	caaaacaaca	gaggggcttg	40680
catggggccag	tgatgataaa	gggctaagaa	tgagaatatata	attaattcaa	ttcctcacac	40740
ctgaggtcta	aaaccaagga	aagggaggggc	caggcgtgga	ggctcacgcc	tgtaatccca	40800
gcactttggg	aggctgaggc	gggcgggatca	caagattagg	agtttgagat	cagcctggcc	40860
aacacagtga	aagcccatct	ctacaaaaaa	tacaagaatt	acccaggtgt	gggtggcacat	40920
gcctgtagtt	agctactctg	gaggctgagg	caggagaatc	acttgaaccc	aggaggcgga	40980
ggttgagggg	agccgagatc	acaccattgc	actccagcct	gggtgacaga	gtaagactct	41040
gtctcaaaaa	aataaaaaaa	ataaaaaaac	agagaaaagg	aggaaactag	atccaggctg	41100

73/122

actagataca	gccttttagag	ttagaaaaaga	tgattttgaca	atctaagccc	acactcagat	41160
tgaatgaaat	tgaaaagcct	ttcaaactaa	aacattttaat	tacaccatct	gctgcagaca	41220
gaactcagac	aactcaaaca	ggtaatgtca	gcgtgggtgtt	ttatatcacc	accctcaaca	41280
cagaataaaaa	atcagctgca	tgtgaagcag	tgactagaat	gaagaaaagg	ctgcttctta	41340
cttccttcta	gtgggttcttt	ccgaaaacat	taataggcac	cagctctatg	catgtcacc	41400
tgcagggaga	catgggggtat	ataactatga	cttactgttc	attcctcaag	gaattcccaa	41460
tcttgtggaa	gattatacac	aatgaggcaa	caaaaactat	ccaataaaaac	cacggaaaag	41520
aagccagtga	caaagaagcc	agtgatgaaa	ggccctgtga	gcagagctga	tggccatttg	41580
gggaagaaaag	accaacatgg	atgggggtga	tcaggggtggc	tccgtgggaa	agctggaaga	41640
gaagtggcag	atctctgagc	tggatgatgg	gccactacca	tctgtatatg	gctaattaaa	41700
gaccatgtgt	ggattttttaa	ttcagctctt	tcgtgtcatt	cctgctatca	gcacagaacc	41760
caatctcaac	tttccagcta	tattgagcta	aacttctcac	ctcatggaat	ttgcagataa	41820
agttcaaaaag	gatccttgcc	ttttcaaaaat	aattttgaaat	ggttgagtag	tcctctgtgt	41880
ctctctcact	gacaccctct	caaggtgtgt	gagcacgtgc	catgctatgg	ctttctccaa	41940
catcaggaaa	tgttttccac	tcagtttccac	cttaatacaa	atgtgttctc	tcttcagaga	42000
aggcaaaaaa	attcatgacc	atctgactgg	ggaaggtcat	ttctaggtaa	agtgtccatc	42060
tttttctgag	gaacacagga	ggaaaatctt	acagaaaaga	gttaacacag	caggcctaag	42120
actgcttttt	aaaataaata	aataaataaa	taaataaata	aataaataaa	taaataaata	42180
aataaatgaa	tgatagggtc	ttctgtattg	gccaggctag	tctcaaatc	ctggcttcaa	42240
gagatcctcc	caccttggtc	tcccacagtg	ttggggattat	agacatgagc	cattgtgctt	42300
ggcccaagac	tgttattctt	aaaaagtcct	ataaaaagca	tgggttaatcc	ttggctggca	42360
cctgggaact	tagatttcag	aagggttccc	accatccaac	ctggaaagag	ggactcactg	42420
tgcctaaatt	attgtgtggt	ttatgctgaa	ctcctgcttt	tcttcaggta	gcgtggaatg	42480
tggatgtgtc	tgggcaaagg	gggcctgcac	gaccagcccc	caataaaaac	cctgggtggt	42540
gggtctctag	tgagtttccc	tggtagacag	catttcacat	gcgttgtcac	agctccttcc	42600
tcggggagtt	aagcacatac	atcctgtgtg	actgcactgg	gagaggatgc	ttggaagctt	42660
gtgcctggct	tcctttggac	ttggcccat	gcacctttcc	ctttgctgat	tgtgctttgt	42720
atcctttcac	tgtaataaat	tacagccgtg	agtaacacc	atgctgagtc	ttccaagtga	42780
accaccagat	ctgagcatgg	tcctgggggc	ccccaacaca	gaaataaatt	ataaaagacc	42840
aaggactggg	catggtggcc	catgccggta	atctcagcgc	tttgggaggc	cgaggcagga	42900
ggaccagtta	agcccaaaaag	ttcaaagtta	cagtgacctta	tgactgcgcc	aatgcactct	42960
aacctgggag	acagagcaag	accctgtccc	caaaaacaata	aactaaacac	atacttctgc	43020
cttccaagtg	tcttaaaatt	caatggaatg	gtagaacat	ttttaaaaca	ctaaatcaaa	43080
agaaaactgg	aaaacaagag	tgcgatggc	caactaaaat	gtctaggaaa	tttctgaaaa	43140
gtaaaaagta	ctcagaacca	gattacctga	gcaaaccata	gccaataaca	agcttgggag	43200
gaggctgtta	tgcagaagga	aatggtaaca	ggtttccagg	aacagacttg	taacagcaga	43260
tagaacagca	gaggtagaac	ctgacaaggt	gattacctgg	ggaactgcag	tctgaatgac	43320
caggactggt	ggacccttcc	cctcacatgg	aatacacagc	ccactcagca	gcacaccaca	43380
gctcttcaac	aatcacagga	ggcacgctac	gcctagtaag	acaggaaaaa	aggaattctc	43440
aaacttcgaa	gatgaacaca	taaagaatca	ccaagttttt	attcagtatg	atgaaacagg	43500
gacactgaat	caacagaaca	caaaccgaag	caaagataat	tactagagca	catagaagaa	43560
attattagat	attcttggga	agacctaaag	ggacattata	aagagcaagc	agttggtatg	43620
tgacgatctt	tgtgatatac	caagaaataa	aaacacagga	tgaagaccag	atagagaata	43680
atgtactat	ttgtgcaaaa	aaggagaaat	ggagaatctg	attcatattt	gcttgtattt	43740
gcatgaagaa	actttggaag	gtacataaag	aactaacaac	aatggttacc	tacttgtaag	43800
gcgagagaag	taagaggaca	ggaatggtgg	gaacaccttt	tgtgtccgga	attggtgggt	43860
tcttgggtctg	acttggagaa	tgaagccgtg	gacctcgcg	gtgagcgtaa	cagttcttaa	43920
aggcggtgtg	tctggagttt	gttccttctg	atgtttggat	gtgttcggag	tttcttcctt	43980
ctgggtgggtt	cgtagtctcg	ctgactcagg	agtgaagctg	cagaccttcg	cggcgagtgt	44040
tacagctctt	aagggggcgc	atctagagtt	gttcgttcct	cctggtgagt	tcgtgggtctc	44100
gctagcttca	ggagtgaagc	tgcagacctt	cgaggtgtgt	gttgcagctc	atatagacag	44160
tgcagaccca	aagagtgagc	agtaataaga	acgcattcca	aacatcaaaa	ggacaaaact	44220
tcagcagcgc	ggaatgcgac	cgcagcacgt	taccactctt	ggctcgggca	gcctgctttt	44280
attctcttat	ctggccacac	ccatatcctg	ctgattgggtc	cattttacag	agagccgact	44340
gctccatttt	acagagaac	gattggtcca	ttttctagag	agctgattgg	tccattttga	44400
cagagtgtg	attggtgcgt	ttacaatccc	tgactagac	acagggtgct	gactggtgta	44460
tttacaatcc	cttagctaga	cataaagggt	ctcaagtc	caccagactc	aggagcccag	44520
ctggcttcac	ccagtggatc	cggcatcagt	gccacaggtg	gagctgcctg	ccagtcccgc	44580
gccctgcgcc	cgcactcctc	agccctctgg	tggctcgatgg	gactgggcgc	cgtggagcag	44640
gggggtggtgc	tgtcaggggag	gctcggggccg	cacaggagcc	caggaggtgg	gggtggctca	44700

74/122

ggcatggcgg	gcccgcaggtc	atgagcgcgtg	ccccgcaggg	aggcagctaa	ggcccagcga	44760
gaaatcgggc	acagcagctg	ctggcccagg	tgctaagccc	ctcactgcct	ggggccgttg	44820
gggcccggctg	gcccggccgct	cccagtgccg	ggcccggcaa	gcccacgccc	accggggaact	44880
cacgctggcc	cgcaagcacc	gcgtacagcc	ccggttccc	cccgcgcctc	tcctccaca	44940
cctccctgca	aagctgaggg	agctggctcc	agccttggcc	agcccagaaa	ggggctccca	45000
cagtgcagcg	gtgggctgaa	gggctcctca	agcgcggcca	gagtgggcac	taaggctgag	45060
gaggcaccga	gagcgagcga	ggactgcccag	cacgctgtca	cctctcactt	tcatttatgc	45120
ctttttaata	cagtctgggt	ttgaacactg	attatcttac	ctatTTTTTT	TTTTTTTTTT	45180
tgagatggag	tcgctctctg	tcgcccagac	tggagtgcag	tggtgccatc	ctggctcact	45240
gcaagctccg	cctcccgggt	tcacaccatt	ctcctgcctc	aacctcctga	gtagctggga	45300
ctacaggcaa	tcgccaccac	gcccagctaa	tttttttttt	tattTTTTTT	ttagtagaag	45360
cggagtttca	ccatgttagc	cagatggctt	caatctcctg	acctcgtgat	ccatccgctt	45420
cggcctccca	aagtgtctgg	attacagacg	tgagccactg	cgccctgcct	atcttaccta	45480
tttcaaaagt	taaacttttaa	gaagtagaaa	cccgtggcca	ggcgtgggtg	ctcacgcctg	45540
taacccagc	actttgggag	gcccagggcg	gcggatcacg	aggtcaggag	atcgagatca	45600
tcctggttaa	cacagtgtaaa	ccccgtcgct	actaaaaata	caaaaaatta	gccgggcgtg	45660
gtggtgggca	ccggcagtc	tcgctactgg	ggaggctgag	gcaggagaat	ggcgtgaacc	45720
tgggaggcag	agcttgagct	gagccgagat	agtgccattg	ccttccagcc	tgggcgacag	45780
agcgagactc	cacctcaaaa	aaaaaaaaaa	aaaatagaga	cccggaaggt	taaaaaatatg	45840
ataatcaata	tttaaaaaaa	ctcaagagat	gggctaaaga	gttgacggaa	caaatctaaa	45900
tattagattg	gtgacctgca	aaaccagccc	aaggaaacatc	ccagaatgca	gcccataaag	45960
ataaagagag	catttccgct	gggcacagtg	gtatggcagg	ggaattgcct	gagtccaaga	46020
gttgaggtc	acattgaacc	acaccattgc	actccaggcc	tgggcaaacac	agcaatactc	46080
tgtctcaaaa	aaaaaaaaaa	ttaaattaaa	aaagacagaa	tatttgagag	aaaaaaatgc	46140
ttatttcaag	aaacatgaaa	gataaatcaa	gatattctaa	ttcccaagta	agaataattc	46200
cagaagcaga	aaatagaata	gaggcaagga	aacactcaaa	acttctccag	tgccatagaa	46260
atgtgtatta	atcttttagaa	tgaacaggac	taccaaattgc	tgagcaggaa	gaacaaaaga	46320
gatccactct	taagccagtg	tgggtgccc	gcgcagtgcc	tcattgcctgt	aatcccagca	46380
ctttgggag	ccgagggcag	tggatcacct	aaggatcagga	gtttgagatc	agtcaggcca	46440
acatggtgaa	accctgtctg	tactaaaaat	acaaacatta	gctgggtatg	gtggtgcaca	46500
tctgtaatcc	caactacttg	ggaggctaag	gcaggagaat	cacttgaaac	caggaggtgg	46560
aggttgtagt	gagccgagat	catgccacac	tcccagcctg	ggtgacagag	caagattcca	46620
tctcaaaaaa	aaaatccact	cctagacaaa	taatagttaa	attttagaac	accaaggaga	46680
aagaaaaaaa	attgtaaaagc	ttcagagaaa	ataaacatta	actacaaaga	aacgagagtc	46740
agacgcgtgc	acttcttcct	agataccagc	agataaagca	atatctccaa	aattcagaag	46800
gttttaacgt	agaatcctat	acccagtcac	gaatattcac	atggaaaagt	gaaataaaaa	46860
acattgttta	aacatgcaag	ggttcagaaa	gtttaccatt	cacagaatcc	ctgaaaacaa	46920
aaccataata	tcacttaagg	actcattaa	aaataaaagc	aaataaaagc	accaatgatg	46980
agtaataaat	cagaaaaatt	tacagtttac	ctaaataact	gtttatgcat	aatgtatgaa	47040
aacccaaaaa	tttaatatgg	gacagaatta	aatcatgat	aagattcttt	tttgctttac	47100
tcattggagag	ttcacataaa	cagattatct	tttaatagca	agagaaaaaa	atgttttagat	47160
atgtgtgaaa	aactaagggt	acaaaaacag	tgcaaatcca	tttatcatca	ggaaaaatcca	47220
aattaaaacc	acagtatcca	ccagaataac	taaaaggtaa	aagacagaaa	ttaccaagag	47280
ttggcaagaa	tgtggagcaa	ccacatatac	ttctggggta	aataagttgg	tgcaaccggt	47340
actgaaaact	gtttgctagt	atctactaaa	accgagcaca	tgacagact	acaaccaagc	47400
agttccactc	ccagatacac	actcaacaga	aatgcacaca	ctcactcaac	aaaagacgtg	47460
tactagagtg	ttcatgtact	tactattcat	aatagtccaa	aatgcaaac	aaccaactgc	47520
caatcaaagt	caaattgtata	tctatattag	ggatatatac	aatggcatat	acacagcaat	47580
gagaatgaaa	tgaaccagct	cggcacagtg	gttcagcct	gtaatctcag	cactttgggc	47640
gggtaaggga	ggcagatcac	ttgaggtcag	aaatttgaga	ctagcctggc	caacacggtt	47700
aaaacctgtc	cccataaaa	acacaaaaat	tagccgggca	tagtgggtgc	aggcctgtaa	47760
ttccagctac	tcggggaggct	gggttgggag	aatcgtttga	accgaaaagc	cggaggtcgc	47820
agtgagcgga	gatcgtgcca	ctgcactcca	gcctggacga	tagagcaaga	ctccgtctca	47880
aaaaaggaaa	tcaaaaatat	aaaataagat	gacaggaata	atccgcaaaa	gatcagtaat	47940
caaaaataat	ataaaatggg	taaagctacc	tattaaaaga	caaagatttc	acacccataa	48000
ggatagctac	tatcaaaaaa	agagagagaa	tagcaaggtg	tagcaaggtg	gtatggaaac	48060
tgaaattctc	acgcattgct	gggtgagaata	taaaatgggt	cagcctctgc	ggaaaacact	48120
atgctgggtc	atcaaaaaat	taaaaataga	agtactactt	gatccaacaa	ttctacttct	48180
gggtatatac	ccaaataact	gaaagcaggg	tcttgaagag	atatttgtac	acccatgatc	48240
atggcagcat	tattcataat	agctatgatg	tggaaaccaac	ataaatatcc	tttgataaat	48300

75/122

atatggataa	gcaaaatgtg	gtgtatacat	tcaatggaat	attaattagc	aataaaaaatg	48360
aagaaaaatc	tgacacatgc	tacaacatgg	atgaaccttg	agggcattac	attaaatgaa	48420
ataagccagt	tataaaaaaga	caaatactat	atgaggctact	atattagata	ctcattgcaag	48480
gtacctaaaa	taggcaaat	catagagaca	aaaagcagaa	tggtggttgc	caggggctgc	48540
ggtaatggat	acagagcttc	aattttgtaa	gatgaaaaaa	ttctggagat	tggttgcata	48600
acaatgtgca	cacacttaac	actggggaac	tgtaaaactta	aaagtagtaa	atggtaaaaa	48660
taaaaataat	aaataataaa	ttttatgtta	ttttaccaca	atatttatta	aaagacaaag	48720
attaactaat	taaacaaaat	ccagccataa	gctaattggt	agagtaacaa	ttaaagaaga	48780
cacagaaaaat	tgaaaatcag	tgactagaaa	aagatattcc	atataaatgc	taacaaaaag	48840
caagtacagc	aatataaaga	gaatgaacaa	aaaaaaaatt	aaataagatg	gctcgtttat	48900
tcccaaaagg	tacaattcac	caagaagata	caagaattgt	gaacctttta	gcacataaaa	48960
cagcttcaaa	aatacaacat	ttaaagaaaa	atataatata	aacatagaaa	tagtacaaaa	49020
accctacaa	gaatcataat	gggagtcttc	aatacaactc	tccatatcaa	caggtcaaac	49080
agagaaaaaa	aataagttaa	ggatgcagaa	aacctgaatt	accatcaata	aacttgagat	49140
taatatagaa	ctgtataccc	aataactaa	gagttcaggg	aacagtcgtg	actgacagtg	49200
gactgcaaat	taactctgtt	ttaatctttg	ttttctcttc	agcactgtgg	cagaaatagag	49260
atcctaaaaa	ccttccagct	acaaaaacatc	tttttaaaaa	tataaaaaaa	tacaaaaata	49320
actctgaaat	caatagaaga	cacatggtga	aacccaaatt	ctagaatata	gggagaataa	49380
aggcattttc	agatattaca	aaaacagaaa	attgatcatt	gctgaagtaa	tttctaaaga	49440
atgtacttga	gggagaagaa	aaatgttcca	aagaaaagta	tctgtgatac	aagaagggaat	49500
ggaaagtga	gaaatggtaa	acaggtagat	aaagctaata	aatggtgacc	tagaaaaata	49560
caaaaaaat	agcaataatg	tctcgttga	agggttgaag	taaaaatata	attaaggcca	49620
aatgtgaggt	aagtggaaatg	aaagaattag	aagtccttgc	cttggtcaca	ggactgatta	49680
aataaatgag	ccaggttttc	cattcaaaca	gttaaaaactt	gaacaaaaata	aactcaaatt	49740
aagtagaag	ataaaaaaca	gaaattaatg	tcatagaaaa	ataaaaaatc	aatagaatta	49800
atcaataaat	cctgggtta	aaaagctggt	tctttgaaag	gattaataaa	ataatcatta	49860
agcaagtctg	atcaaaaaaa	aagagaaaa	gtacaaaaaa	aagtactgta	tcagaaagag	49920
aacatacaga	tacatacaga	tatgtaaagag	tctgttttct	tacaccagaa	tactatatac	49980
aacattatgc	tagcatatat	taaatttcaa	taatgttaat	gattttctag	gaaaacagaa	50040
aatattaaat	ttactttgaa	gaaacagaaa	aactgagaaa	aataaatgat	catgaaaaaa	50100
atgaaaagg	aattaaatac	tgatattaac	tgcttaaaaa	acaccagcag	cagcccaggc	50160
agtctgcagt	caagttctgc	caaacttgag	ggaacagata	attcttctat	tccagagcat	50220
agaaaatgat	ggaaaagtttc	ccaattta	cagagaggac	agcctgatcc	ttgttatgaa	50280
cacagataaa	aatggggtaa	actatatgcc	aaactcagat	acaaaaaccc	taaaataagat	50340
gctagcttat	tgatgtgaac	aatccaaaag	tgcattttaa	attagcccag	ggtttttagag	50400
aaagaaaatc	tagcaatgtg	accaccactt	atgttaacaa	ttttaagacg	aaaatctaca	50460
tgatcatatc	aatgcatgct	acacaaaagc	atttgggcaa	aaaacccaac	acccaccctt	50520
gactttttta	actcttagta	attaggcata	aacagaaatg	tacttaattg	gatagaatac	50580
actcggtgaa	gatacagagg	gaatgctccc	taaaaccaag	cccaagacaa	agattcctat	50640
ttaacctcaa	tagtcaacac	tgacgcgaga	gtaatctatg	gaagacaagg	aaaaaagtaa	50700
aaacatgaga	gacatctgtt	gtttaacaga	caataagatc	acctacttgg	aagaggcaaa	50760
cgaatcaagc	gaaaaactat	taaaactgag	acaggcttta	gtatggaggc	tcagcttcag	50820
ctgtagtttg	ggctaccaaa	ttcaactcgc	ttgtctggag	agttaatcct	gcaaagctaa	50880
tttctgttga	ggatttagga	ttgacaagcc	tgtgtctctc	cctcctcccc	catcttcaac	50940
actgaaataa	cacggtgttt	ggaactggat	aacagaatct	tccaaaaaca	aaaattgtcc	51000
tgaagggctg	acttgtgccc	ttactcaaaa	aacactttat	ctgctgcctg	cagctcctac	51060
agttgtctgt	ggataagcct	gccaaccagc	tcggcgtaat	tcttctctga	gagggcaagg	51120
aagagcactt	tcacaggaaa	attttttttc	gaactgtatg	ccgcttatta	cataaactta	51180
cgtgctggca	aatggagctc	cagcaaaata	agatattcag	agtcaaaactt	ccttaggaaa	51240
aaaaaaaaaa	aaaagcaagc	acataaacct	aatttccttg	catgggcact	gggggaaggag	51300
gtcgttactt	ccgcacgccc	gcaggtccgc	accaccggga	aaccacggg	caccgcgcgc	51360
tgccccggg	ccttcagggt	gcactgcgc	gcggcgcccc	agctgaccg	ggatgcgcag	51420
ccctagccct	ccccctgtca	ccccggccag	gaaggcgccg	gagcgcggcg	gacgcggagg	51480
gcgaagggct	tctcggctct	ctgcaccacg	cagcaccccc	aaggcacaac	agggagggtg	51540
cgggagggtc	ccgagaccca	ggagccgggg	cggggcgtgc	ccgcgcacct	gtccccactgc	51600
ggcgaggcct	gggggtcgct	ccagggccgc	agctgtcggg	agccacctgg	ctctcagtc	51660
cggggccctg	ggacaaccct	cgggcccggg	ggggaggagg	cggccacctg	ccgctgccac	51720
ctgcccgcac	gggtcccaccg	ctccggggcg	ggcaggacag	gccaggacgt	ccctcctggg	51780
ctggggacag	gacacgcgac	gaggggaccg	gggccccgc	ggcgaagacg	cagcacgcct	51840
tcccagaaa	gcagtcctgt	gccccacga	cggactgcgc	gacccccgcg	ctcgcgccgc	51900

76/122

catcccttca	gaccacgcgg	ctgaggcgca	aagagccggc	cggcggggcg	gctggcgggc	51960
cggctagtag	tcaccggccc	cgctgggtca	gcgcgcgcgc	aacccccagc	ggccacgggt	52020
ccggggcgctc	actgatgctc	aggagagggga	cccgcgctcc	gccggcgccct	ccagccatcg	52080
ccgccagggg	gcgagcgcgga	gccgcgcggg	gctcgctggg	agatgtagta	cccggaccgc	52140
cgctcgcc	gtcctccttc	agccggcgcc	cggggggccc	ctctctccca	gctctcagtg	52200
tctcatctcc	ctatctgctc	atcctctggg	cgcacataat	cgatgtttgg	gcgtcccaag	52260
ccagatgtgg	accccatttc	cgcactctac	actggagggt	ttctaagggt	ggtgcccggg	52320
ccagcagctt	cagcctcatc	tgggaacttg	agaāaatgca	gattctccgt	cccaccagc	52380
ctattcggtt	tttctcgac	taaaaccatg	aaggtggggc	ccagcagtc	acattctcgc	52440
aagcccgctca	agtgattctg	aggcgccctc	cagtttgaga	gctatgctca	cggcctcacc	52500
tccgccccgc	aaggagcccg	gtcttgccctg	tggcgctagc	cgcacacgga	cacctcatcc	52560
tgcggggccc	gccccccgc	tgcacctca	ccgcccacg	cctcctccgg	gatgcagcgg	52620
aggcgccctgg	aagtcggcaa	ggtcaacatc	cccctcagca	tcttccctac	cctcacggct	52680
cctcctccag	gggtgcctca	tggccagggg	ttagaaagag	ccactgtgtt	tcttgacatg	52740
gaagtggcct	aagaccttaa	tgaāaactgc	aggagtggaa	tgacagaacc	tttggtcata	52800
cttgagggcg	tgaagctcaa	atgaggagga	aggaaaggat	ccagggagaa	taaccaaccc	52860
tggcaagtgt	tggcgcccag	gtagaggggc	gagcctaggc	tagcggttct	cgaccagggc	52920
cgggtgttggc	cctcctcgcc	gccccgcgta	catttgggga	ggtctggaga	catttttggt	52980
tgtcatgatg	cgggagttgc	tactgttgcc	taagtgggta	gacacgaggg	tgctcctcaa	53040
catcctacct	gaaggacagg	actgccccac	aaggaagaa	gatccggccc	caaataagaa	53100
accctgggct	ggtcagcaac	aacccctttg	ttctgagaag	agaggaggaa	agaataaaag	53160
aagtgggggtg	aagttttggg	ttggtagagg	aaacttgaag	acattttcac	tggaaaggaa	53220
gagaggaaga	ggagggagat	gtctgtgaag	acgagcaaac	cgggtgacag	ctgatttcct	53280
catattgaag	taatgagctc	tagttataat	aaattcctaa	taaaaaccca	gtttatccct	53340
gcaataaaat	tgtctttttt	ttttaaatat	actgcttgat	tctgtttgct	aatattttat	53400
ttacaggctt	tgcattgata	tgcāaaaatg	agatgggcaa	taattttctt	tttgaatgtc	53460
taatgttgtt	tggtttcaga	atcaatgtta	tgctcacatc	ataaaaaatt	tggaaaccgag	53520
gcaggaggag	tgcttgaggc	cagaagttcg	agaccagtc	aggaaacaca	gtgagacccc	53580
cccattctcta	caaaaaaaa	aaaagaaaaa	aaaatgggca	tgtttgcttt	ttccttttac	53640
tctgaacaat	ttaaggagca	ttāaaattat	ctattctttg	aggtttgatc	atttcccagt	53700
taaaaatggt	cctcccagcc	tgatgttttc	tttggggagg	gtāaatcttt	taaggctaga	53760
aaagtttctt	ctgtggcaat	tttattattt	acatttttaa	aattattcta	gagttaattt	53820
tgataaagca	tgtatttctt	aaaacaaatt	atcctttttt	tccagatgtt	caagtgtatt	53880
tgcataaagt	tgaggaaagt	agtcttttgt	gaatctttta	acttctccca	aatatcttat	53940
tttgtgtatt	tttgcttctt	tattttgtta	acttttaaaa	gtgtattttt	ttttcaaaga	54000
atcagctctt	aggtttatgt	ttttggttat	actggagctt	ttttcttctt	ctttttaaaa	54060
tatttttctt	cctttatttt	ttagacgtat	tttgcataaa	cgtaatcgga	agaaggtaaa	54120
ttagaatctt	ttgttactat	tgtgttttta	tttctcttta	tttctctgaa	gtcctgtctt	54180
ataaatagta	ccatgttatt	tgtgcataaa	tattcatttg	tcttatattc	ttgggaattt	54240
tcccacttca	tcataaaatg	accttccttg	tctcatttaa	tgtgttcaaa	ctttgccctg	54300
aatttaactt	tgtctgatat	tttaccatcc	tgtgaattt	tgtttggttac	cccaaacac	54360
ctttgctggt	ttcgtctttt	ctgaaccctt	tatttttaggt	aatcccttga	attagagcac	54420
taagttttgc	tttgtgatta	aatctgaaaa	tctttatctt	gccatagatg	agttgagccc	54480
tattcatgtg	acagctatat	tatgctgttt	catagccctt	ttggctcctt	tttcactctt	54540
gcattgcata	ttttgtgttt	attgtgtttt	gtgtttcttc	tgataatttg	gaaggtttgt	54600
atttttattc	agggagttgc	cttataatca	tactccgcaa	tacacatcgt	cctcagtttc	54660
ttcagactgt	ctgttaactc	cctattctga	ataaaaaatga	cattgttaatt	tccctctttt	54720
ttctttaccc	cttttcttct	cctcacctaa	tgtāaatgat	tttatccttc	tttagtattt	54780
gcttttttaa	ttaactacat	ttataaatat	ctttatcact	tgatttttaa	atcagctttg	54840
aatgagatat	ttggattcct	agatataaaa	gatgttaatt	ataccatttc	cacgttagta	54900
gggttatataa	atcatatatt	ctgctgtgta	accataatcc	cacgtttggt	ttagttccac	54960
tcctacagtt	aaaagattca	gaagtattat	taacagttat	tttgccatag	ttttttcccc	55020
aacccatttt	gtggtaagtt	atgatcctgc	tttagtttct	taagaataat	ttatagagca	55080
gagtgtggtt	gctcacgttt	gtaatcccag	cactttggga	gacaagaggt	agaaggatcg	55140
cttgaagcca	gcagtccaag	accaccctga	gcaactatgt	gagaccttgt	ctctacaaaa	55200
aatttttaaaa	tttagccaga	cgtagtggcg	tgtgcctata	gtcccagcta	ctcaggaggc	55260
tgaggcaaga	ggattgctag	agcccagaag	tttgaggctg	cagtgcacctc	tgatttgtgcc	55320
actgcacccc	agtctgggca	agaaagttag	aacctatctc	tttaaaataa	caataataac	55380
ttatgaaaaat	tatattccct	gagtttttca	tgtttaaaaa	tatttgtttgc	ctttatcctg	55440
taaaagtttg	agtataaatt	cttgggttat	actttattta	ttgaagaatg	tataagtatt	55500

77/122

gtcttctaga	attgagtgtt	gctgtaatga	aaccagaagt	cagcctgggt	tatttttctt	55560
cagaaatgag	gtaattgccg	gccggacacc	gtggctcatg	cctgtaatcc	caacactttg	55620
ggaggccgag	acaggtggat	cacgaggtca	ggagattgag	accatcctgg	ctaactgtgt	55680
gaaaccccg	ctctactaaa	agtacaaaaa	gttagctggg	catgggtggg	gacgcctgta	55740
atcccagcta	cccgggaggg	tgaggcagga	gaatggcggt	aacctggggg	gaggagcctt	55800
cagagagctg	agatcgcgcc	actgcactcc	agcctggggc	acagagtggg	actccgtctc	55860
aaaaaaacaa	aaaaaaaaaca	aagaagtga	gtaattgcca	tgatgctcca	agaattatct	55920
ctttgtctat	gaaatccaga	aatctcactg	ttatacattt	tgggaattatt	attctggggc	55980
aatatttctt	gggacacaa	agattgactc	tatagattta	attttttttt	tttttttgag	56040
acagagtctc	actgcaatct	cagcttactg	caacctctgc	ctcacgggtt	caagcaattc	56100
tcttgcctca	gcctcccaag	tagctgggac	tacaggcgcg	tggcaccatg	cctggcta	56160
ttttgtcttt	ttagtagaga	cagggtttca	ccatgttggc	caggctgggt	ttgaacgcct	56220
aacctcaagt	gatccacctg	cctcagcctc	ccaaagtgtt	gggattacag	gcgtgagcca	56280
ccatgcccag	cctcaattcc	tctttctatc	tggttaattt	tctgaagtgt	aaaacatttg	56340
ttctaatacg	tattttcagt	gttcttctaa	gatgtgtaaa	gcacctattt	cccaggctcag	56400
ccccactctt	gctagtgagc	tcggctgggt	cttcacaga	gctctgggtt	tctcctgctt	56460
aatctcaagt	acctctgtca	gcctccacct	ggtttatgat	ttggagtgtt	ttgggttttg	56520
ttttttgttt	ttgacagagt	cttactctgt	caccaggtct	ggagagcagt	ggcataatct	56580
cagctcactg	caacctctgt	ctcccaggtt	tgagcgatct	tcttgcctca	gcctactgag	56640
tagctgggat	tacaggcgcg	tgccaccaca	ccgggcta	ttttgtattt	ttagtagaga	56700
tgggttttca	ccatgttggc	cagggtgggt	tgaaactcct	gacctcaggt	aatccacctg	56760
cctcagcctc	ccaaagtgtt	gagattacag	gcgtgagcca	ccgcgcctgg	catggtttgg	56820
agttttaatc	tgtagtttta	ataaagatag	tgcttatgtt	tgtgtttctt	atatttcttg	56880
gtactcttgg	gtaatttgta	agatcccat	atctacacaa	gaagtccatt	ttcaattctt	56940
ttcttcagac	tgtttatttt	attttatttt	attttatttt	tatgtttgag	atggagtctc	57000
gctgtgtcac	ttctggaggc	tggagtgcag	tggcgcgatc	tcagggtcact	gcaacctccg	57060
tctcccgggt	tcaagcaatt	ctcctgcctc	agcctcccga	gtagctggga	ttacaggcac	57120
ctgccacttt	ttaatttttt	tagagacaga	gtctcgcttt	gttgaccagg	ctggagtgcg	57180
gtggtgcaat	catggctgac	tataacctcc	aaatcctggg	ctcaagtgat	cctcctgctt	57240
cagcctcctg	agtagctggg	actacaggca	catgccacca	tgcccagtta	atttttaatt	57300
ttttgtagag	acagggtctc	catatgttgc	ccaggctggc	ctcctactcc	tggcctcaag	57360
taactcctct	acctcagcct	cccaaattac	taggattata	agcatgagcc	accatgcccc	57420
gccttgttct	actactttaa	tttcatatgt	taggtgacca	tgtaattgat	catccaaacc	57480
aggatactgt	aagaatgaaa	gaggctgaca	gtagtatgat	gctgggacta	gcattgtgca	57540
ctgagattat	ttctgggaaa	gcaggagata	cggtcaccct	acttatagtg	tgcttgtctt	57600
tggattgttg	aatttggagt	ttctatttgc	aggcttattt	caactgggca	gccttgatcc	57660
gccctgcccc	gcaatgtcac	cgttctctcc	accgggtctc	tgggaccctt	tcagtacta	57720
tacttagctc	agttccccac	cctcccactc	cctaaaagcg	taaccaggaa	tcttgcctca	57780
ggctcactgc	cgtcttccgt	gggctgtttc	agttcctatt	accagagatc	aaactcccag	57840
cattccctac	ctgattccag	acttggagtc	cagagcttta	acctcttcag	gccaaactcc	57900
cactttgcat	ttctgtccct	atatcttagt	ccatggagat	acatttcatg	tctttgagtc	57960
tacttacaaa	gtaaattttg	ctgtttttta	attttttttt	tgagatggag	tcttgccttg	58020
tcacccaggc	tgtggtgcaa	tgacgccatc	tcggctcact	gcaacctccg	cctcctgggt	58080
tcaagcgatt	catctgcctc	agcctcccaa	gtagctgtga	ttacagacag	gcaccaccac	58140
gcccagctaa	ttttttttat	cttttagtag	agacagggtt	tcaccatgtt	ggccaggctg	58200
gtcttgaatt	cctgacctcg	tgatctgccc	atctcggcct	cccaaagtgc	tgagattaca	58260
ggcgtgagcc	actgtgcccc	gccaattttg	ctttttttat	atttcattgc	tatatgttta	58320
gaggataagt	ttacagtgtt	atatgcattc	ccaaatatta	gacaaaaaaa	atctccaaaa	58380
aattagaaag	aaaatccaaa	aaatctcaaa	aaataccaaa	aagcaacaat	ctcacagacc	58440
atactcactg	accccccaata	aaataaaaat	agaaattaac	cacaacttaa	caaaataaag	58500
tactcaagtc	agagaggaaa	gaggaaataa	acatacaaat	tacaaagtct	aggcggtggc	58560
tcacgcctgt	aatcccagca	ctttgggagg	ccaaggcggg	cagatcacaa	ggtcaggaat	58620
tcgagaccag	cctggccaat	atggtgaaac	cccggtttcca	ctaaaaatac	aaaaattagc	58680
caggcatagt	gatgtgtgcc	tgtaatccag	ccacttggga	ggctgaggca	ggagaatcac	58740
tgaacccagg	gagacgaaga	ttgcagttag	ccaaaatcgt	gccactgcac	ttcggcctgg	58800
gtgacaaaagc	gagactccat	ctcaaaaaaa	aaaaaattac	aaactcttta	gatagaaatt	58860
ttgggtgtttt	tttttgagac	ggagtctcac	tctgtcgcag	aggctggagt	gcagtgggac	58920
tatgtcagct	caccgcaacc	tccatctcct	ggattcaagc	aattctcctg	tctcagcctc	58980
ccaagtagct	aggattacag	gcgcccacca	ccagaccag	ctagttttta	tatttttagt	59040
agagatgggt	tttcaccatg	ttggccaggc	tggtctcaaa	ctcctgacct	caagtgatcc	59100

78/122

acctgcttca	gcctcccaaa	gtgctcagat	tacaggcggtg	agccaccgca	ccccacctag	59160
atagaaattt	caacatgagg	ccgggcacaa	tggctcacgc	ctgtaatctc	agcacttcag	59220
gaggctgagg	cgtgggagga	tcacttgggc	ccaggagtgc	aggaccagca	tgggtgacag	59280
agacagaccc	tgtctctatt	tatttgaaaa	aaaaaaaaaa	aaagagagag	agaaagaaat	59340
ttcaacatga	aaagtatctc	tcaaaccctt	cgagatggtg	gcaaaaagcg	actcaaagga	59400
aaatgatta	ctgtgtgtga	atttgcttga	aaataagaaa	gaggccgggt	gtggtggcta	59460
acacctgtaa	tcccaacact	ctgggagtc	gaatcaagt	gatcatgagg	tcaggagatc	59520
gagaccatcc	tggctaacat	ggtgaaaccc	tgtctctact	aaaaatacaa	aaaattagct	59580
aggcgcggtg	gctcatgcct	gtaatcccag	cactttggga	ggctgaggca	ggtggatcac	59640
ctgaggtcag	gggtttgaga	ccagcctggc	ctacatgggt	aaacctcgct	tcttctacaa	59700
atacaaaaat	tagctgggag	tgggtgggtg	tgcctgtaat	cccagctact	cagaggctga	59760
ggcaggagaa	tcgcttgaac	ccgggagggc	gaggttgagg	tgagccgaga	tcgcaccact	59820
acactccagc	ctgggcaaca	gcctgggtga	cacagtgaga	ctccatctca	aaaaatacaa	59880
aaaattagct	gggtgtggtg	gcctgcgcct	gtagtcacag	ctaccgggga	ggctgaggca	59940
ggagaatgga	gtgaacctgg	gaggaggagc	tgcagtgag	ccgagatccc	accactgcac	60000
tccagcctgg	gcgacagagc	aagactcttg	tctcaaaaaa	aagaaaaaaa	aaggaaaaaa	60060
gaacctgat	aataaagaaa	ccaaatgttc	aactctcaaa	gctcggacac	tttaaagaaa	60120
taattaataa	aggcagaagt	taaagggagg	atgataaagc	aatttttttt	gttgggtttt	60180
ttgagatgga	gtcttgctct	gtcaccagag	ctggagtgc	gtgatgcat	cttggctcac	60240
tgaacctct	gcctcccggt	ttcaagcaat	tctcctgct	cagcctcctg	agtagctggt	60300
actacaggtg	gcgcgccact	ggcccagcta	atttttgtat	ttttattaga	gacgggggtt	60360
caccatattt	gttaggctgg	tctcaaactc	ctgatctcag	gtaatctgcc	cacctcggcc	60420
tctcaaagt	ctgggattac	aggcaggcgc	caccgcgcct	ggcctaaagc	aaaatatggg	60480
ttctgtgcaa	aaggtcaata	aaaagagcaa	acgtttacaa	actggagcca	gcacccattc	60540
agctcagtg	gtctggagaa	aaaacaatct	cgcttcagaa	ttcatgatta	cgcagccctt	60600
tttgcttct	aaaaatccta	ctatgttgct	gttgaccatt	ctctctcttt	ctctctctct	60660
tgttttctct	ccagaaaagc	tattcagaca	ttctctctct	tcctcaaacc	tccaacactt	60720
cctcctccat	ccttagcctc	agctgctgac	ctcacttcta	atcattgaga	aaccaggaga	60780
agcatttaag	agtgaacctc	cgctccccc	cacggggcaa	accaccacc	cacagaattg	60840
tgccccaatt	ctgcgtcctc	tcctctcacc	atggatggac	ggtccaggct	ccgagccaaa	60900
gccaggcctc	ccctggagct	ctggatccac	cacctgcagc	ttctcaggca	gggccccagc	60960
agctccccctg	ctcccttgta	ccatcaatcc	ctccccctac	tgggtcactc	ccaacaatat	61020
atatatttga	tgatgtttct	cccatgtggt	aaaactcact	agcctctctc	ctccccagc	61080
tactatctta	tttggtttct	tccattctct	gcaaaaactc	tcaaagcatt	gtgtctatgt	61140
gctgactcca	tttatcttct	cccgttctct	gctgagtcct	tcccacagac	tctcacccca	61200
gttactccat	gaaatgacct	ctgcactgcc	acatccaatg	gtgaatgttc	agttcttaat	61260
tttattcagt	ctttcagcag	catttgacct	ggcgcgatc	tcctctctct	taaaaaact	61320
tttctcagcc	aggcgtgatg	gctcacacct	gtaattccaa	cactttggga	ggccaaggcg	61380
ggaggatcat	gagagcccag	gagttcaaga	tcagcctggg	caacatggca	agaccctatc	61440
tctacaaaaa	ctaaaaagta	gccagtgtga	tggcatgcac	ctgtagtccc	atctacttag	61500
gaggctgagg	cagtaggatg	acttgagcct	gggaaatcaa	ggctgcagt	agccatgatt	61560
gcaccactgc	actccagcct	gagtgcagc	gagaccctgt	ctcaaaaaga	caaaatagga	61620
aacttttctc	agcatattcc	tctgattctc	ctgctgcttc	tgtctgcaca	gattcagtct	61680
cctttgccc	ttcttctctc	tcctctctgat	ctcttgacct	tgaagtgcc	cagagtacag	61740
tctttttttt	tttttttgag	acgcagtctc	gtctgtcacc	caagctggag	tgcaatggcg	61800
aggctctcagc	tcattgcaacc	tctgctctct	gggttcaagc	gattctcctg	cctcagcctc	61860
ccaagtagcc	aggactacag	gcacatgcca	ccatgcccag	caaattggtg	tatttttagt	61920
agagacaggg	ttttactata	ttggccacgc	tggctcctcaa	ctcctgaact	cgtgaaccac	61980
ccgcctcggc	ctcccaaagt	gctgagatta	caggcatgag	ccaccacacc	cggcccagag	62040
tacagtcttt	agacggcctc	tctacctata	cttgctcccc	tcataaactc	ctcctgcctc	62100
atggcttttaa	ataccatccc	tagactgatg	actcccatat	ttctcttttt	tttttgga	62160
cggagtctcg	ctcagctccc	caggctggag	tgcagtgccg	cgatctcggc	tcactgcaag	62220
ctccacctgc	caagttcaca	ccattctcct	acctcagcct	ctccagtagc	tgggactaca	62280
ggcaccgcgc	accacgcctg	gctaattttt	ttgtattttt	agtagagatg	gggtttcacc	62340
atgttagcca	ggatggctct	gatctcctga	cctcgtgatc	cgcccatctc	ggcctcccaa	62400
agtgctggga	ttatagggtg	gagccaccgt	gcccagccga	tgactcccat	atttctatct	62460
cttgctgtgt	gggagttctc	ctcagaactc	catactcata	aatccaactc	tcataaatag	62520
tatctcaaat	gggcaatatg	ctcaaaagtc	aatctctact	tttctcccta	aacttgcttt	62580
cctgcagtct	ccaccatctt	aatgtccaat	ctaacattag	gaggcaaaaa	ctttgaagtc	62640
attcttgact	cttctctatt	acacacccta	tccaatcttt	ctgcagatcc	agtcgacccc	62700

79/122

caaataccagt	tagctctcat	catctccct	gttaccctct	ggtccaggcc	atcttccct	62760
ctcacctgaa	tcactgcagc	attctcctca	ctggctctct	tgggtctgtt	ttcactccac	62820
cttagcatag	tctccacaga	gcagtcagag	ggatcctttt	aaagtgtaat	tcccatcctg	62880
tccctgctct	gctcaaaacc	ctgtcgtgat	tcccgtttta	atctgtcaga	ttaaaagcca	62940
gagtccttcc	agtgccttac	atgatctgcc	tattatcacc	tcccacttct	ttccccttgc	63000
tcactccact	ccagctctgc	agctgtcctt	tctgtttcct	gaacagccca	gattttgctt	63060
ctttaagaac	tttgtatttg	ctgtcccttc	tgtctggaat	gtttttccag	gaagtcacct	63120
ggctctctcc	tgcacttctt	tcctgaccac	catgtttaaa	aatcactcaa	acacacttca	63180
ggccggacat	ggtggctcac	gcctgtaatc	ccagcacttt	gggaggccaa	ggtgggtgga	63240
tcacctgagg	tcaggagttc	gagaccagcc	tggccaacat	ggtgaaactt	cgtctctact	63300
acaaatacaa	atagtagcca	ggtgtagtgg	cacacacctg	taatctcagc	tactcaggag	63360
gctgaggcag	gagaatcgct	tgaaccagag	aggcagagga	ggtgcagtga	gccaaagatca	63420
cgccacaaca	ccccagcctg	ggtgacagag	caagacccca	tctcaaaaaa	aaaaaaagaa	63480
aaaaaaatca	cacaaacaca	cttctcttca	tattcctttt	ccaagtttta	tttttctcca	63540
gaatacttta	cattgtttta	atggaagttc	tccgtttccc	cccaactaga	atggatactt	63600
cctgcaggta	ggcactctag	tcctccctac	caagtactaa	ccaggctcaa	ccctgcttag	63660
cttctgagag	caggggagat	caggcctgtt	cagggtggta	tggcccagga	attttgattc	63720
tgttttattc	attgctgttc	tgttgattct	cttttgttcc	tcctcctagt	gctgagaaca	63780
ctacttgtac	ataataagca	ttcaataaat	atgtgttgaa	tgaatgactt	gttgaatgaa	63840
ttaactcagc	aaatgcagga	ctgggtctac	attagaaaaa	ttttcaaggt	cattctctgt	63900
tgtcgtaaca	cattaaagaga	ggaaaatttt	gtactctaaa	tcatttgata	aaatacatat	63960
tgattttctgt	tttcaaaaac	tcttagtggt	tgggagaggt	ggctcacatc	tataatccca	64020
gcattttggg	aggacgaggt	gggaggatca	cttgagggtca	ggagtttgag	accagcctgg	64080
ccatcatggt	gaaaccctat	ctctactgaa	aatagaaaaa	ttagccgggt	gtggtggcgc	64140
atgcctgtag	tcccagctac	ctgggaggct	gaggcaggag	aatggcttga	accgggagg	64200
cggagggtgc	agtgcagcaa	gatcatgcca	ttgactcca	gcctgggtta	cagagtggag	64260
ctccatctca	aaagaaaact	cttagtgagt	ttaggaatcc	aaggaagacc	ctcaactaa	64320
atagataatc	tagctaccag	aagccttcag	taaaccttaa	cactccatgg	tgaacatta	64380
gaaacattcc	tactaaaaga	caggctaaga	atgcctgcaa	tcctcacggc	tagtccaaga	64440
agtcaaaaag	aagaaatgag	cgctgattta	aaaaaatata	caaacaaaaa	actaccgatg	64500
cagaggctgg	cagcaaggac	tgaaggactg	tacagtactt	gcctggagca	ggcggatggc	64560
cacacccctg	cgaagcctgc	tcagctggct	gggggacgct	ccagtgtgtg	agtggcagga	64620
tgcagggtac	ttcctctgcc	agggagtgtc	actggggaga	tcctcccca	ctcacacttt	64680
ggcagctggg	gctttggaat	gtgacttagc	ttctgtcaaa	gggtcaatcc	accctttgat	64740
atatgatgca	aaggcgaaca	tatgatgcaa	aggtgagaga	acagcccaaa	ttaggacttt	64800
taccacagct	gtggaggtgg	acagcgacag	tgggtgggcc	tggccagact	tttcatgctc	64860
aaaggtgggt	gttgttcttc	ctacttctgt	tccctccagg	gcttcctttg	cctgtgtgct	64920
gaactgtctt	cttttaactt	tttttaactt	tttttaactt	ttaattgttt	taattaaaac	64980
aaattttgaa	aactgtctga	acctgctttt	gaacctgtct	atgatttgaa	tgtttgtccc	65040
ctgccaaact	gattttgaaa	cttaatctcc	aaagtggcaa	tattgagatg	gggctttaag	65100
cagtgcactg	atcatgagag	ctctgacctc	atgagtggat	taatggatta	atgagtgtgc	65160
atgggagtgg	catcagtggc	tttataagag	gaagaattaa	gacctgagct	agcatggtcg	65220
cccttcacc	atttgatata	ttacactgcc	taggggtctt	gcagagagtc	cccaccaaca	65280
agaaggctct	caccagatac	agctcctcaa	ccttgtaact	ctcagcctct	gtaactgtaa	65340
gaaataaatg	ccttttcttt	atgaattacc	cagtttcaga	tattctgtta	taaacaatag	65400
aaaacgaact	aaggcaaaact	ctcatgattc	tactgccatg	ccattccaat	aaactccctt	65460
tatgtcttaag	agagccagag	ttggccaggc	gtggtgactc	acgcctgtaa	ttccagcact	65520
ttgggaggcc	gaggcagggt	gatcacaagg	tcaggagatc	gagaccatcc	tggctaaccac	65580
ggtgaaaccc	cgtctctact	aaaaatacaa	aaaaattagc	tgggcgtggg	agtgggtgcc	65640
tgtagtccca	gctactcggg	aggctgaagc	aggaggagaa	tggcgtggac	ccaggaggcg	65700
cgctctgcag	tgagtgcaga	tcgtgccact	gcactccagc	ctgggtgaca	gaatgagact	65760
ccgtctcaaa	aaaaaagaga	gccagagtgt	atttctgttg	cttgcaacca	agaaatctgg	65820
ctgggtgcact	gaagtttcca	taaataatag	caattttaaag	actctttcca	agccaggcaa	65880
tgcctagcct	tgtgtagtcc	ttgtggtaat	acattcattc	attcatttgt	tcaaccaact	65940
gtgtctccaga	gactaagaat	acaaaaatgg	gggcccgggt	tgggtggctca	cacctataat	66000
cttagcactt	tgggaggccg	aggcaggtag	atcacctgag	gtcaggagtt	cgagaccaac	66060
ctggccaaaa	gggtgaaacc	cctactctac	taaaaataca	aaaaattagc	tgggggtggg	66120
ggcggacacc	tgtaatccca	gctactcgtg	agactgaggc	aggagaatca	cttgaacccg	66180
ggaggcagag	gttgagtgga	gccgagatcg	caccactgca	ctccagcctg	ggcaacaaga	66240
gcgaaactcc	acctcgaaaa	aaaaaaaaaa	aaaaaaagag	ggccggggct	gggcgcagtg	66300

80/122

gctcacgcct	gtaatcccag	cactctggga	ggccaaggca	ggagaattac	gaggtcagca	66360
gatcgagacc	agcctgacca	acatggtgaa	accccatctc	tactaaaaat	acaaaaatta	66420
tccgggctg	gtggcgacac	cctctagtcc	cagctacttg	ggaggctgag	gcaggagaat	66480
cgcttgaacc	cgggaggcag	aggttgagct	gagccgaaat	catgccactg	cactccagcc	66540
tgggtgacag	agtgagactc	cgtctcaaaa	aaaaataaaa	aaaaaaaaaa	gaattcaaaa	66600
attgtagagt	tatagtgtgc	ttctagttaa	gttgagagga	catctgtcct	tcaagggaag	66660
ctagaatcta	taccctgagt	ccttactgaa	atcaatccag	cagtcaaaac	atgggaccaa	66720
cgatcacagc	agtaagatag	gaagagcacc	tttgtacatt	tagctcatgt	tgagataagc	66780
cactgacaga	gctgaaggaa	gctcacagtt	ctgggttcca	tcctttggca	tttaaaaaaga	66840
aaagtgctaa	gaaaattcgg	ttggtcacgg	tgggtcacgc	ctgtaatccc	aacactttga	66900
gaggccaagg	caggcagatc	acgaggtcag	gagttcgaaa	ccagcctggc	caacatggtg	66960
aaaccccgct	tctactaaaa	acagaaaaat	tagccgggca	tgggtggcgca	tgccataaat	67020
cccagctact	caggaggctg	aggcaggaga	attgcttgaa	cccgggaggg	ggagggttgca	67080
gcgagtgaga	gcaggccact	gcactccagc	ctgggagaca	gagcaagact	ctgtctcaaa	67140
aaaaaaaaag	aaaaaaaaagaa	agaaaaggaaa	aaaagaaaag	aaaaaaaaaaga	aaaaagaaaa	67200
ttcaggccag	gccaggccctg	gtgggtcaca	cctgtaatcc	caacactttg	ggagggtgaa	67260
gcgagacggt	gccttagccc	aggagttaga	gaccagcctg	agcaacatag	cgagaccctg	67320
tctctataaa	aaaaaatttt	tttttggcca	gacgcagtgg	ctcacgcctg	taatcccgagc	67380
actttgggag	gccgaggcag	gtggatcacg	agggtcaggag	atggagacca	tcctggctaa	67440
cacggtgaaa	ccccatctct	actaaaaaat	acaaaaaat	aaccgggcgt	ggtggcgggc	67500
gcctgtagtc	ccagctactc	gggaggctga	ggcaggagaa	tggcgtaaac	ccgggaggcg	67560
gagcttgca	tgagccgaga	ttgcgccact	gcactccaga	ctgggagaga	gtgagactcc	67620
gtctcaaaaa	aaaaaaaaaa	aaaaaaaaaat	taattgtcag	gtgtgctggc	atgcagctgt	67680
agtcctagct	actcgggagg	ctgaggttaag	agatcgctt	gagccaggga	gttcaaggct	67740
gcagtataag	tgctctcac	tctaccctgg	gtgacaatga	gacctctct	caaaaagaaa	67800
gaaaaaagg	aaagaagaaa	agaaagaaa	aaagagaaga	aaggaaggaa	gaaagaaaga	67860
aaaagaaaa	gaaggaagga	agaagaaaaa	aaaagaaaaga	aagaaaagag	agagaagttc	67920
aaagaccaaa	gggtcaggat	cccaaaatag	tttttatggt	ttattttatt	atttacttat	67980
ttatttttga	gacagtatgg	ctctgtcgcc	caggctggag	tgagtgatg	cgattgcggc	68040
tcactgcagc	ctccaaactg	ggctcagggt	gccctcccac	ctcagcctcc	cgagtagctg	68100
ggaccacagg	cgcggtgccac	catgccagc	taatttttta	attctttgta	gagatgaggt	68160
ctctatatgc	tgcccagggt	ggtctcgagc	tcctgggctt	aagccatcca	cccgcctggg	68220
cctcccaaag	tgctgggatt	acagaagtga	gccaccgcgc	ctaactcggg	ggtttgtttg	68280
tttattgacg	gggtctcgct	gctgccagg	ctggagtgcc	agtggctgtt	cacagggtgca	68340
gtcctggagc	attgcatcag	ctcttgggct	ctagcgatcc	tccagagtag	ctgcagctgg	68400
gattccaggc	gcgccaccgc	gcggggctca	gaatgggttt	ttatattgag	ggttatgctg	68460
ccacctagag	gatatatgta	gtaccgaact	gtgtgcgcag	ggaggctgag	ggtgcagtga	68520
gccaagatga	tgccagggca	ctccagcgtg	ggtgacagag	caagatttca	tctcaaaaaa	68580
aaaaaaaaaa	aaaaaaaaaa	aagaattgaa	agtaaggctc	tgaagagata	tttgtgcctg	68640
tatgggtcata	gcagtattaa	ctttgaccca	ctagctaaaa	cacaaaagca	acatgtgtct	68700
gtcagcaggt	gaacggataa	acaaaatgtg	gtatatatgt	acaattgaat	attattcagc	68760
ctttaaaaag	gaataaaaag	ctggatgcgg	gggtcacgc	ctgtaatcct	aacactttgg	68820
gagactgagg	tgggtggatc	acccgagggt	aggagttaga	gaacagcctg	gccaacatgg	68880
tgaaacttca	tctctactaa	aaatactaaa	attagccggg	catggtggca	cttgtctgta	68940
atccaagcta	ctggggaggc	taaggcagga	gaattgcttg	aactcaggag	ccggagggtg	69000
cagtgaagta	agatggcacc	actgcactcc	agcctgggca	acagagttag	actccatctc	69060
aaaacaaaaca	aacaaaaaat	tattatttcc	aaagaaaaca	gacctgggt	ccatttccca	69120
gcccacacct	gatgttgact	cacaacacac	agcctgggtt	gctatgagcc	tgcttcattt	69180
aattgtcacc	ttacttcac	atcacctcca	agtcctggaa	taactctttg	ctgacctttg	69240
tgtgctgagc	catctccatg	tcgctcaacg	tgagctccct	ctcactgcac	tgagtcaata	69300
gccagacgtg	gtctgactgc	agggctcatcc	ttgttggtct	aggctgactc	gggcatagca	69360
gggtgctctg	agacctcacc	gcatataggc	tttgcccca	ataaactcta	tataatatcc	69420
atattatgtg	gtctgggtgt	gtgtagcttt	gcactgtctt	ctcgtgacag	tgccctcaac	69480
ctctttccca	ggatttcctc	ctctacctcc	tcaagtccca	ctgctctgca	aagaccaaaa	69540
gctgcagagt	cccagctccc	tcctttacac	cccacgacgc	agcctcctct	ctcagaaccc	69600
tttaaacaga	gtcttttact	gcagatccca	agaacagcca	cacccctctc	tccccccac	69660
tccagacaca	cccaggtaat	tatagcacc	agggttaacta	tgtagatgga	gtccctggaa	69720
catgtggata	gtgccccctg	ggagtatgca	aaagcaacat	tgctggcacc	tgagagaaac	69780
agggtgacat	ccaggaatca	gagcatgggc	ctctgggagg	tagggatgtg	gccaggcagg	69840
ctgccaacaaa	ttggtagagc	aaggccacag	gatctttctg	accttccttc	caaacagagg	69900

81/122

ctcctgtact	ggtgatccct	gtgttgattg	accactccct	tcctgggggt	cgtgggtctct	69960
gtcccagttg	cccggacttc	tgtgagtgtc	ctactgaggt	ccttttcatg	agaagcatgc	70020
tgctcctcca	cctgctggga	gcaagagtga	caacttcaat	actataatag	cagtggcata	70080
cagagaagaa	gaaagatgaa	gtggcaagaa	aaacaggctt	ccaagcagga	gtttttctat	70140
aaaaacaaaa	acgtttacaa	gcaaactttt	tataaagggt	tagatagtaa	atatttttagg	70200
ctttgagagc	cacatagact	tgtttgcagg	gactcaatgt	cgctattgta	gtttgaaagc	70260
agccatcagg	gttatgtaaa	tgagtgagtc	tgattttggt	tcagcaaaat	tttatttacc	70320
aaaacagaca	atgagtgggc	tggatttggc	ccatgatcct	tagtttgcca	actcctgctt	70380
tgggctcacc	cagatctgat	tttgaattct	ggctctgcta	ctggtttagct	gcaggagctt	70440
ggaaggctct	ctgagcctgt	ttcctcatct	gtaaaaat	agcaataatt	tctaacactc	70500
aagagtgtta	cctcacgcct	gtaatcccag	cactttggag	gctgaggcag	gcggatcacc	70560
tgaggtcaga	agttcaagac	cagcgtggcc	aacgtggcaa	aacctgtct	ctactaaaaa	70620
atacaaaaag	tagccgggca	tgggtggcgcg	catctgtaat	cccagctact	tgggaggctg	70680
aggcagggat	actgctagaa	cctgggaggt	ggagcgtgca	gtgagtggag	atcacacctc	70740
cacactccag	cctggccgac	agagcagagc	tccatctcaa	aaaaaaaaaa	aaaaagagtg	70800
ttagaagggt	ttgagataat	gaataaaaaga	tgcttctgtg	atactaaagta	ttcaacaact	70860
gatagctgca	ttggtcta	tataacagtt	tagaagcgat	tgagtcaaca	aatgctggat	70920
ttgtcaggga	ggacttccta	tcaggaggta	gatcttgggc	tgagtcctga	agcaaagata	70980
ggcattggat	agaggagttg	agagaacacc	ctaggactgt	tattattatt	attcgacacg	71040
gagttctctg	ctctgtcacc	caggctggag	tgcagtggcg	cgatctcggc	tcactgcaac	71100
ctctgcctcc	caggttcaag	cgattctcct	gcctcctaag	tagctgagac	tacagggtgtg	71160
tgccaccaca	cccggcta	ttttatattt	ttagtagaga	cagagtttca	ccatgttggc	71220
catgctggtc	tcgaactcct	gacttcaggt	gatccacccg	cctcagcctc	ccaaagtgtc	71280
ggaataacag	atgtgagcca	ccgcacccag	cccagaacca	tttttcaatc	cttggctctg	71340
ccttttatta	gctgcaagat	ctcaggcaat	ttattttaacc	tctccaaaga	ctcattttct	71400
cattcacaaa	atgaggcaaa	taataatata	tactatccca	ggttgtcatg	agaattaaat	71460
gcaacatgac	atttaaatgaa	atgagaagtc	ccttggacat	taactggcta	aagtatgtgc	71520
tcgacaagga	tatcatttta	ggtggatact	tagcatctca	gaactgatgc	tcacaatgga	71580
aatatcattga	aacgcattaa	aattcatttt	aaatgattgt	aggtagtgtg	gcaattgaaa	71640
gaagaagaca	agaggactga	ttataatgct	tcaggctcac	tagtctcctt	ttaggaggga	71700
aaaacaattt	caagttaaat	tttaggctct	agatttttac	ccctgctgct	cattagaatc	71760
accagatttg	atgaaatcag	agcccatctg	aggctgtgtt	tttcatctcc	agaatgagag	71820
ctgttgtggg	gattaagttt	ttgaaaaagt	acatctaaca	ggtgatcgaa	aatgatagt	71880
atattattgc	agttaggttc	attattgttg	ttattattat	actgaaagag	gcttcagttt	71940
tctgatccat	aaagtgaggg	aattgcatga	gaccattgct	aagattcctt	ctagctctgt	72000
ttttttgttt	ttgtttttta	gacagagtct	ctgtcgccca	ggctggagt	caatggcatg	72060
atcttggctc	actgcaacct	ccgcctcccg	ggttcaaatg	atcctcctgt	ctcagcctcc	72120
gaagttagct	ggactacagg	cacacaccac	catgcccaag	taacttttat	atttttaata	72180
gaggtggggt	ttcaccat	tggtcaggct	gggtccaaac	tcctgacctc	aggtgatcca	72240
cccgcctcgg	cctcccaaca	tgtctgggatt	acaggcatga	gccactgtgc	ccaacccctt	72300
ctagctttct	tgatcactga	ttctagggtt	ctctgctgaa	atataattga	gacatcctgg	72360
ataaaaagatc	atgcaagagc	tcaccaatag	gtattaataa	ttgattctgg	aggcttagct	72420
actcctgatg	gattagacat	gactcaactg	cctctcttat	gtgtacaaca	caacaacaca	72480
accaagaaaag	gttattctgg	cattccattt	attcagttta	tttacagccc	ttacttccag	72540
cagcacgtta	aagatatggc	cagggccggg	tgcagtggct	caagtctgta	atcccaggac	72600
tttgggaggc	caaggtgggc	ggatcacaag	gtcaggaggt	tgagaatctg	gcaattcttc	72660
agacttagaa	gcaaccagct	cgataaacaca	gtcttgtgtg	ggctctccct	ctgtccctcc	72720
ctcgcttccc	tcatttctca	tccctgcccc	tgagactgtg	caccttcaca	tagccctgcc	72780
atgagacctt	catctcaggc	tttgctttct	ggggttaactg	aggctaaaca	ctgagtggcc	72840
ctaaaagagg	attgggattt	ggaagttaga	ttattcacca	gagaacagac	tttgctgatg	72900
atcaggccca	ggttgtaatt	gttgaaaaaa	agagaggatg	catagtctta	tctcatctcc	72960
tagtcaaaagt	ataaacatg	ataaataaga	gtcaaatcct	gagatgtgaa	ttggggacat	73020
ttgagtgggt	aaccttgaga	agcttgcacc	ttcagacccc	tcaatacccc	tgctccccag	73080
agaaggctgg	acattgacct	cagcacaggc	aggagccctg	caagatgccca	tttgtcctac	73140
taaagatgga	cccctccact	ctgtttctag	gtaaaataacc	aaagtcaagt	ctccacacag	73200
cctgagcaag	aaagtccag	cctgctacag	gagaaaaatac	cacactggcc	aaaggattca	73260
ctagccctgg	ccactgtgtg	tgggaggaac	caggggaatca	tgtgtgggag	tcaattgtga	73320
agctgttgga	ctgggggtgg	ggtggaatat	aagcctggcc	ctggggagtt	tttcccgttt	73380
gagggccttt	accacaact	caagatccag	tgtatagaca	ggagatccca	gagctagtcc	73440
taacagatgg	tcaggattga	acttggccta	gagtaaaatg	aggaggatag	tgccagaact	73500

82/122

ttctcaacat	actattgagg	aagagggtcag	aaggcttaag	gaggtagtgt	aactggaaaag	73560
gggtcctgat	ccagacccca	ggagagggtt	cttggacctt	gcataagaaa	gagttcagaga	73620
cgagtcacc	cagtaaagt	aaagcaattt	tattaaagaa	gaaacagaaa	aatggctact	73680
ccatagagca	gcgacatggg	ctgcttaact	gagtgctctt	atgattattt	cttgatttcta	73740
tgctaaacaa	aggggtggatt	atthgtgagg	tttccaggaa	aggggagagg	atthcccgaga	73800
actgatggat	ccccccactt	ttagaccata	tagagtaact	tcctgacgtt	gccatggcgt	73860
ttgtaaactg	tcattggccct	ggagggaatg	tcttttagca	tgtaaatgta	ttataatgtg	73920
tataatgagc	agtgaggacg	gccagagggtc	gctttcatca	ccatcttggg	tttgggtggg	73980
tttggccggc	ttctttatca	catcctgttt	tatgagcagg	gtctttatga	cctataaact	74040
ctcctgccga	cctcctatct	cctcctgtga	ctaagaatgc	agcctagcag	gtctcagcct	74100
cattttacca	tggagtgcgt	ctgattccaa	tgctctgcag	agcaggaatg	ttggaattga	74160
attactatgc	aagacctgag	aagccattgg	aggacacagc	cttcattagg	acactggcat	74220
ctgtgacagg	ctgggtgggtg	gtaattgtct	gttgccaggt	gtggactgtg	ggagatgcta	74280
ctactgtaag	atatgacaag	gtttctcttc	aaacagggtg	atccgcttct	tattctctaa	74340
ttccaagtag	cacccccccg	ctttctcttc	ctttctcttc	tttctgattt	tactacatgc	74400
ccaggcatgc	tacggcccca	gctcacattc	ctttccttat	ttaaaaatgg	actggggctg	74460
ggcgcggtgg	ctcatgcctg	taatcccagc	actttggggg	gccgagggcg	gcggtatcatg	74520
aggtcaggag	atcgagacca	tcctgggtta	cacgggtgaa	ccccgtctct	actaaaaatg	74580
caaaaaacatt	agccaggcgt	ggttgcagggt	gcctgcagtc	ccagcggtct	aggaggctga	74640
ggcaggagaa	tggcgtgaac	ctgggagggtg	gaggttgcaa	tgagccgaga	ttgtgccact	74700
gcactccagc	ctgggtgaca	gagcgagact	ccgtctcaaa	aaaaaaaaaa	aaaaaaaaaa	74760
tagctgggca	tggtggcgcg	tgctgtaat	accagctact	ctggaggctg	aggcaagaga	74820
atcgcttgaa	cccagtaggc	ggaagtgtga	gtgagccgag	atcttgacac	tgcaactccag	74880
cctgggtgaca	gagttagact	ctgtctcaaa	aaaaaaaaaa	agaaaaaa	agacagaaag	74940
aaagagcaca	gacagagtca	caggtatttg	cagtaggaag	ctgtcagggt	agagtgcacg	75000
gaaatagaaa	gtatatttta	cacttacagc	acatctctgt	ttgattagcc	acatttaaaa	75060
tactgaatag	caacgtgtgg	ctatttagta	ttcactaaaa	tcttggacag	tgcaagtcta	75120
aagaatcctt	gatccgtccg	gcatgggtgg	tcacgccttt	aatcccagca	ctttggggagg	75180
ccaaggtgga	agggtcactt	aaggtcagga	gttcgagacc	agcctggcca	acatgggtgaa	75240
acctcgcttc	tactaataat	acaaaaaa	ttagccgggc	atgggtgggtg	atgcctgtaa	75300
tcccagggtac	ttgggagggt	gaggcaggag	aatagcttga	atccaggagg	cgctgcagtg	75360
agccgagatc	atgccatgcc	actactgcac	tccagcctgg	gcaacagagt	gagactgtct	75420
caaaaaaa	aaaaaaattg	ttgggcgtgg	tggtcacgc	ctgtaatccc	agcactttgg	75480
gaggctgagg	gggggtggatc	acctgggttc	tgaggttcga	gaccagcctg	gccaacatgg	75540
tgaaacccca	tctctactaa	aaatacaaaa	attagctggg	cgtgggtgggtg	ggcacctgaa	75600
atctcagcta	ctcaggaggc	tgaggcagga	gaatttcttg	aaccaggagg	gcagagggtg	75660
cagttagcca	agatcgcgcc	tctgcactcc	atcctgggtg	gcagagcaag	actatgtctc	75720
aaaaaaaaaa	aaaaaaatc	ttgattgtct	ggacattctg	cagaacatca	tatggagaca	75780
ctatgttgac	gacatcatgc	tgattgttaag	caagaaatgg	caagtgttcc	agaaacacag	75840
tcaagacaca	tacatgccag	aaggtgagat	ataaactcta	ctaagattca	gtggcctgcc	75900
acactgggtga	catttttaaa	cctgctagat	gtttgtgtag	aaaaggattt	aaccttgccc	75960
aaagagggtg	ctggcctttg	tccccagcta	ctggacataa	tctcttttaa	ctcttgaaat	76020
atcattcctg	atagaagtat	ttttgttttg	actaggggcc	ttgggcccagc	cagatagcaa	76080
caatgtgatc	tggtttgggg	gctttggatc	aggtggcatc	agtgtgacct	cctgagtggtc	76140
tagagactag	aatcaaccac	atgggcagac	aaccagctt	acatgatgga	attccaataa	76200
agactttgga	cacaagggtc	tggttaagct	ttcctgggtg	gcaatgctct	atactgggaa	76260
acctattctg	actccatagg	gagaggacaa	tggtatattc	tcatttggtg	cctccctggg	76320
ctttgcccta	tgcatTTTTT	ccttgtctga	ttattattat	tattatgaga	tggaatctcg	76380
ctctgtcacc	caggctggag	tgcatgggaa	tgatctcaac	tcactgcaac	ctctgcctcc	76440
ccggttcaag	cgattttctt	gtctcggtct	cccagtagc	tggtgactaca	gatgcatacc	76500
accacaccg	gctaattttt	ttgtattttt	agtagagacg	gggtttcacg	ttagccagga	76560
tggtctcgat	ctcctgacct	catgttccgc	ctgctcggc	ctctcaaagt	gtaggaata	76620
catgtgtgag	ccaccgcgcc	cagccccctt	ggctgattat	taaagtgtat	ccttgagctg	76680
tagtaaat	taaccgtgaa	tataacagct	tttagtgagt	tttgtgagca	cttctagcaa	76740
attatcaaac	ctaaggatag	ccttggggac	ccctgaactt	gcagttgggtg	tcagaaataa	76800
gggtgctcat	gtgtgtacca	tgccctctaa	tttgtagtt	aattaacttt	cacaacttta	76860
ttattaccgc	ttacactcaa	tggtttattca	cattttatcca	cataccactt	attctagtgc	76920
cttgcatcaa	agactttcta	tctcatgtac	tttattctgc	ttgaagtaaa	tccttttagga	76980
tattcttttt	tttttttaaa	ctttgcacat	acatactttt	atthttttatt	tatttttaat	77040
tttgttattt	ttgtgggtac	gtagtagata	tatgtattta	tggagtacat	gagatgtttt	77100

83/122

gatacaggca	tgcaatgtga	aataagcaca	tcatggagaa	tgggggtatcc	atcctctcaa	77160
gcaattttatc	cttcaagtta	caaacaatcc	aattacactc	tttaagttat	tttaaaatgt	77220
acattttaatt	ttgtattgac	tagagtcact	ctgttggtgct	atcaaataata	atTTTTTTTT	77280
tttttgagac	agagtctcac	tcagtggccc	agactgaaag	tgcaagtggca	caagctcggc	77340
tcacttcaat	ctctgcctcc	ctgggtcaag	cgaatctcct	gcctcagcct	cccacatagc	77400
tgggattaca	ggcacacacc	accatgcccc	gctaattttt	atattttttt	agtagagacg	77460
ggttttcgcc	atgttggcca	ggctgggtctt	gaactcctgg	cctcaaataga	tctgaccacc	77520
tcagcctccc	aaagtgttag	gattacaggc	atgagccacc	acacctggcc	aaaatagaat	77580
attcttttagt	gaggtctgct	ggtgacaatt	tttttctttt	ttttgagact	gagtcctcgt	77640
gttgctcagct	tgggctggag	tgcaatagca	cgatctcagc	tcactgcaac	ctccacctcc	77700
cggattccag	caattctcct	gcctcagcct	cccaagtagc	tgagagatta	caggcaccca	77760
ccaccacacg	cggctaattt	ttgtattttt	agtagaaatg	gggggttcacc	gtgttggcca	77820
ggctgggtctc	gaactcctga	cctcaggtga	tcacccacc	ttggcctccc	aaagtgtctg	77880
gattacaagc	atgagccacc	acgcacagcc	aattttttcc	gtttttgtct	gaaatcttat	77940
tttgtgtcat	ctttgaaata	tattttttagt	ggatataaaa	ttgttgggtg	atagttatta	78000
tcattattat	tattattttg	agacaggttc	tcactctggt	gcctatgctg	gggtgtagta	78060
atgtgatctc	ggttcactgc	agacttgacc	tcctagggct	caggtgatct	tcccacctca	78120
gcctccctag	tagctgggac	tacagatgca	tgccaccata	cccaactaat	ttttctattt	78180
tttgtagaga	tgaggctttg	ccacatttcc	caggctgggtc	tctaactcct	gagctctagc	78240
aatccaccca	ccttggcctt	acaaagtgc	ggggcatgac	tagccagcag	ttacttttta	78300
tagcatattg	aatatttaat	atgaatcttc	tggcatccac	tgtaactgtt	taaaaaatca	78360
gctgtttact	tggcactctt	tttttttttt	ttttttttga	gacagagtct	tgccctgtcg	78420
cccaggctgg	agtgcagtg	cgtgatcttg	gctcactgca	agctctgcct	cccgggttca	78480
gcctattctc	ctgcctcagc	ctccggagta	gctgggacta	aaggcgcccg	ccaccacgcc	78540
cggctgattt	ttttgtattt	ttcgtagagt	tggggtttca	ccgtgttagc	caggatggtc	78600
tcgatctcct	gacctcgtga	tctgtccgcc	tcggcctccc	aaagtgtctg	gattataggc	78660
gtgagccacc	gcgcccagcc	tctttttttt	tttttttttag	acggagtctt	actctgtcat	78720
ctaggctggg	gtacagtggc	gtgatctcag	ctcagtgcaa	cctccacctc	ctgcctcagc	78780
ctgcccataa	gctgggatta	caggtgcgta	ccatcacgcc	cggctaattt	ttgtattttc	78840
agtagagatg	gggtttcacc	atgttagaca	ggctgggtctc	gaactcctgg	cctcaagtga	78900
tctgcctgcc	ccagcctccc	aaagattaca	ggcatgagcc	accgcacccg	gccaagtagc	78960
actcctttga	aggtaatctg	cttcccctac	ccctagcaat	ttttaacaat	ttttcttcat	79020
ttttatttcc	tgaagttttg	ttattaataa	tctgtgtgca	gattttctttg	tattttctttt	79080
gtttgcagtt	catagtgatt	cttgaattag	tgtgttggtt	tctgttatca	ccacagga	79140
attgtcagcc	gttagctttt	caaataattc	cttgctaaat	tctctcttct	cccctttcgg	79200
tacaattgat	ttgattaaaa	ctaaaaccag	ggccgggtgc	agtgactcat	gcctgtaatc	79260
ccaacacttt	gagaggctga	ggcagggtga	tcacctaagc	tcaggagttc	aagaccagcc	79320
tggccaatat	agtgaacccc	cgtctctact	aaaaatacaa	aaattaccag	gcatgggtggc	79380
acacatttgt	agtcaggagg	ctgaggcagg	agaattgctt	gaatccagga	ggtggagggtt	79440
gcagtgaagc	gagatcccac	cactgcagtc	tggcctgggc	gacagagtga	gatgagaatc	79500
tgtctcgaaa	aaaaaaagta	tgaatgtttg	ataaactata	tttgttagaa	tgtttgttgt	79560
agaatactat	tcattgattt	ttaaacaatg	ttagattaaa	ccattcactg	gattttgtgat	79620
aattaaactta	ctgatttttac	ctcactgatt	tgttgtaatt	aatacaactg	gtataaaaaag	79680
actgtgacga	ggccgggcat	ggtggctccc	gcctataatc	ccagcacttt	gggaggctga	79740
ggcaggcgga	tcacctgagg	tcaggagttc	aagaccagcc	tgaccaacat	ggtgaaaccc	79800
catctttact	aaaaatacaa	aattagccgg	tcgtgggtgg	gcatgcctgt	aatcccagct	79860
cttcggggagg	ctgtggcagg	agaatcactt	gaacccggga	ggtggagggtt	gcagtgcagc	79920
gatatcgcg	cattgacatc	cagcctgggc	aacaagagcg	aaactccgtc	taaaaaaaaa	79980
aaagaaaaaa	aacacataaa	acaaaacaac	actgtgacgg	ttcccaaaaa	ttaggagcat	80040
aattaaagga	actcctgata	aaaatttaatt	ttatcttaca	tgtaaaactaa	aatgacttta	80100
tgaagttaat	tcagaaatac	aatgcagggt	attagtttgc	cacagctgcg	tattcagcct	80160
aatgtaatat	tcttgttatt	ttttaaattct	tcttttaact	ttactcatat	gtggatcatc	80220
aaatttcaaa	agattaaatg	acaatactct	tagcagcaag	cttccctaag	catataaaca	80280
ttttaatggg	tgatgattca	gaaggtagcc	gaagaatatg	tactgccaga	tatcattcac	80340
ccccatatac	ctgcccgcga	gacatcccat	tttgggaccc	tggataaatg	tgtgggtgga	80400
gagaaagata	ggagaaagtg	gtataagcaa	atggcttttg	agtctgattg	acagcgattg	80460
aaatcctgtc	tctacctctt	aacagcctca	tgatcctaca	taagttaccc	cgatcctcag	80520
ggccacatct	gtaaattggg	ggttgcgatg	gcagccatct	cacagggtct	cttttcgggg	80580
aagggcagga	attatggatt	aagtgcagct	gtaattgtaa	agcacttaat	acaaggaggg	80640
cgcataataa	gtacttcata	aataatgacg	gccattatca	tgactgaggt	gtatgcagct	80700

84/122

gtcggggatt	acggcgactt	cagaattttct	ggtggggcagg	gctcaaaggc	agcaaatacac	80760
actggaagtc	gaggtgaggc	actgcttctg	cacagactgc	ttagctggag	agaatgagga	80820
aggcttagag	gagatttaga	ggaacttaga	gtcctccgcc	tccaactctg	tgggatctgc	80880
tcccgtgcc	gagacattca	ggggattttct	cgcactctcc	cctcccctac	gtccctcccg	80940
ccccatccaa	ctaaccacac	aacacataca	aaatagcccc	tgcgaggttc	tgcacgctgg	81000
aagggaacag	gagaaggcg	ctgcgctttc	ttgctgatgc	cctgtacttg	ggccccctggt	81060
agacacagcc	acttgctccc	tcagcctgca	gagaaatccc	acgtagaccg	cgccccgggtc	81120
cttggcttca	gccaatctcc	ctttgggtggg	ggtgggatgc	acgatccaag	gttttatttg	81180
ctacagacag	cgggggtgtgg	tccgccaaga	acacagattg	gctcccagg	gcatctcgga	81240
tccctgggtg	ggcgccgctc	agcctccccg	tgcaggcccc	gccgaggcca	ggaggaagcg	81300
gccagaccgc	gtccattcgg	cgccagctca	ctccggacgt	ccggagcctc	tgccagcgct	81360
gcttccgtcc	agtgcgcctg	gacgcgctgt	ccttaactgg	agaaaggctt	caccttgaaa	81420
tccaggcttc	atccctagtt	agcgtgtgac	cttgagcagt	tgactttatt	tttcagtgcc	81480
tagttttcca	gataccagga	ctgactccaa	ggactattac	tcactctggag	ggttttagcac	81540
agtaccgtcg	catagtaaat	ttccatgtca	gtttttggtta	cctttcatgc	acttgcaaac	81600
atgccatgct	ctcgaaccaa	ataggcacat	cttttttttt	ttttttttta	aggagtcttc	81660
ctctcgccca	ggctggagtg	cagtggcgcg	atccttggtc	actgcaacct	ccacctcccg	81720
tgttcgagat	tctcctgcct	cagcctcctg	attagctggg	actacaggca	tgccacgacg	81780
cccagttaat	ttttgtattt	ttagtagaga	cgggggtttcg	ccatcctggc	caggctgggtc	81840
taactctcga	ctccagggtga	tctgactgcc	tcagcctctc	aaagtgttgg	gattacaggc	81900
ataagccact	gcatctggcc	agaaatgaaa	taagttaaate	ttttaacctg	ctctaacaat	81960
atagtgaata	gaccatatta	ttattagagc	aggttaaggg	atttgcctat	ttcgggttct	82020
agttatagtc	ttaaacttgg	acattcctgt	agaaagtaaa	aagtttcctc	ttcaaagtcc	82080
cccttcttgt	taaagaatac	atcataagtg	ttagaagtaa	tagtttattt	taaagactaa	82140
ctttcttcaa	gcctccttgc	tttgtgctaa	taactccttg	tttaagcccta	tcctatgtaa	82200
ctgttgagca	tgctcacagg	caggttccag	ttcacagcct	atgccccctc	cttatttggga	82260
aatgttattg	cttccttaaa	cctttcggta	agcaacttcc	tctccttctt	cgttcttctt	82320
tgactttacc	tatttagaaa	gttttaggct	attagcaaat	cggctatcag	tttaagagtg	82380
tgaggtcccg	ctccagccaa	tggatgcagg	acatagcagt	gaggacgacc	caaatgcgta	82440
agggataaat	atgtttgctt	ttcctttgtt	cagggtgtgct	ctcgacatcg	ttccatctgc	82500
gattgagcac	cctttctgca	gaaagtaaa	attgccttgc	tggagatctt	ttgtctccgt	82560
gctgactttt	cttcgtggca	ccgattatct	atttctaaca	attttggtat	ttctaacatt	82620
ctgaacaate	ttgggctagt	tgtctcttct	gggcctgttt	ccccatccgt	cacatgataa	82680
acttcatttg	tttaaaaacc	ccagcgaaca	tttattgagt	tactattacc	ttcctgcctt	82740
ccccaacccc	aaccccagg	agcagttaca	acctcagcgg	ctgagcgcac	tcgcccgggtg	82800
ttaagaagca	ccaaagacag	ggaggcttga	ttgattttgc	tttgggagta	gagggtcaga	82860
agattcacag	gaaaatggca	tttgagcaag	gatgattcac	tggagctagc	ttttaaatac	82920
tggcgaggct	tttatgttgc	agtcctttac	aaagttagag	attcgagggg	actgcactcc	82980
gaaataagcc	cgcttccctt	tttcatctgc	taatgatcca	gggagctgct	ggttcgcctt	83040
gcggcaggtt	gtgccttttc	ctaatacagg	ttctgcatcg	cctcgaaccc	gcaggccgtg	83100
gcgggttctc	ctgaggaagc	agggactggg	gtgcagggtg	aagctgctcg	tgcgggccag	83160
cgcctgtgag	caaaaactcaa	acggaggagc	aggaggggtc	gagctggagc	gtggcagggt	83220
tgaccctgcc	ttttagaagg	gcacaatttg	aagggatacc	agggggccga	agccggggac	83280
ctaaggcccc	ccccgttcca	gctgctggga	gggctcccg	cccaggagg	tagttttgca	83340
gagactgggt	ctgcagcgct	ccaccggggg	ccggcgacag	acgccacaaa	acagctgcag	83400
gaacggtggc	tcgctccagg	cacccagggc	ccgggaaaga	ggcgcgggta	gcacgcgcgg	83460
gtcacgtggg	cgatgcgggc	gtgcgcccct	gcacccgcgg	gagggggatg	gggaaaagg	83520
gcggggccgg	cgcttgacct	cccgtgaagc	ctagcgcggg	gaaggaccgg	aactccgggc	83580
gggcggcttg	ttgataatat	ggcggtggga	gctgcctggg	catcccagg	aggcggtggg	83640
gcccactccc	ggaagaagg	tcccttttcg	cgctagtgc	gcggccccctc	tggaccggga	83700
agtcggggcc	ggttgctgaa	tgaggggagc	cgggcccctc	ccgcgccagt	ccccccgcac	83760
cctccgtccc	gacccgggcc	ccgcatgctc	cttcttcggg	cggaaaaggta	gctgaggggg	83820
cgccggcggg	gagtcaggcc	gggcctcagg	ggcgggcggtg	gggcagggtg	gcctgcgagg	83880
gctttcccca	aggcggcagc	aaggccttca	gcgagcctcg	acctcgccgc	agatgcccc	83940
tgagtgcctt	gctctgctcc	gggactcttc	tgggaggag	aaggtggcct	tcttgccgga	84000
ggtcagagga	gtattgtcgc	gctggttcag	aagcgattgc	taaagcccat	agaagttcct	84060
gcctgtttgg	ttaagaacag	ttcttaggtg	ggggttagtt	tttttggtgt	tctttgagga	84120
ccgtggatca	agatcaagga	aatctcttta	gaaccttatt	atggaagtct	gaagtttcca	84180
aatgttgagg	gttttatgtc	taaaagcaac	acgtgaaaaa	attgttttct	tcacccagtg	84240
ctgtcttcca	atttcctctt	tggggggagg	ggtagtact	gctgttacta	aaataaaatt	84300

85/122

acttattgct	aaagttcccc	aacaggaaga	ccactacttt	tgatgacttt	ggcaagtttg	84360
ctaactactg	gaaccctaac	ttacaaacga	actacttaca	tttttgattt	ccagttgtat	84420
tacctgcccc	atgtttacgt	agaaacagct	taattttgat	tctgggtaac	gttggtgcac	84480
ttcattaaaa	atacatatcc	gaagtgaagca	agtatgggtc	tgtggacagc	agtgattttt	84540
cctgtcaatt	cctgttgctt	cagataaaat	gtaccagaca	gaggccgggc	gcggtggctc	84600
acgcctgtaa	tcacagcact	ttgggagggt	tggcggtgg	atcacctgag	atcgggagtt	84660
caagaccagc	ctgaccaaca	tggagaaacc	ccgtgtctac	taaaaataca	aaattagcca	84720
gggtggtggc	gcatgcctgt	aatgccagct	acttgggagg	ctgaagcagg	agaatcgctt	84780
gaacctggga	ggcgagggtt	gcggtgagcc	gagatagcac	cattgcactc	cagcctgggc	84840
aaaaagagcg	aaactccgtc	tcaaaaaaaa	agtaccagac	agaaatgggt	tttgttttct	84900
ttttttgttt	tgagacggag	tttcgctctt	gttgcccagg	ctcgagtga	atggcgcgat	84960
ctcagtctcg	gctcactgca	acctctgtct	cccaggttta	atcgattctc	ctgcctcagc	85020
ctcccaagta	gctgggatta	cccatgcccc	accatgcccc	gctaattttt	gtatttttag	85080
tagaaaacggg	gcttcacat	gttaggtgg	tcttgaaccc	ctgacctcaa	gtgggcctcc	85140
caactcggcc	tcaccaagt	ccaggattac	aggcatgagc	caccgcggcc	agccagaaat	85200
gggttttggg	aaaagcacta	aacaaaatcg	aacttgggtt	catatgacag	ctctgctgct	85260
aactgtaaca	ggggcagacc	agttaaccta	cttttctgtc	ttctgtcagc	tgagaattag	85320
atgattccca	aaggccatt	gaactctgaa	tgactttaaa	tacttcttct	taagtgggtg	85380
cacggttttg	gtaactgatg	ccagggtgat	aatgcatgaa	agtgtttaat	gaatgaaacc	85440
ggtaaaatag	taggaggaag	ctttatttgt	aaggcagggg	tatacctaata	agctctctaa	85500
tttattggta	ttgaagtgg	taacttttgt	tttttaagg	ggggaaaaca	ttctaagaat	85560
aatgaggcaa	actgcatatt	gcacaagaga	ctgttgtctc	tattcaacaa	ataccttttg	85620
agtgtccaga	gtctgccagg	tgctgtgcta	ggccctcacg	attgagtagt	gaaccagaga	85680
atgtccctgc	acctatggag	cttattgtct	actggggtag	acagataata	aataagcaaa	85740
caaactctct	ctcttctccc	tttcgctcca	tgtaagtgtg	tgtgtatagg	tgtatactta	85800
caagttgagt	aaagtgttat	gaaagattaa	gaggagaaat	gcatttttgt	tagatgttag	85860
aggactcagc	aggtgacctt	gaaacttaga	gctgaaggat	cagtaggagg	taactagaga	85920
ggccagggaa	tcgcatgttc	aaaggccagg	aggcaagaaa	gagcatgggt	cccttcaaga	85980
gaggaagaa	ggctactgtg	actggagcat	agatgtaggc	aagtgttggg	tgattgagag	86040
ctctacgggc	catggttagg	ttttattcct	aatgccgaga	tgccaaacat	ggtggttcat	86100
atctgtaatc	ccagtatttt	aggaggccga	ggcaggaata	tagcttgaac	ccaggagttc	86160
aagaccagcc	tgagcaacat	gagacctgta	caaaacattt	aaaaaattgc	tggggtatgat	86220
ggtgcacacc	tgtggtcccc	gctactcagg	aggctgaggg	agaaggatca	cttgagccta	86280
ggaggtggag	gctacaatga	gcatattttg	agtcactaca	ctccagcctg	gatgacaaag	86340
tgagaccatg	tgtcaaacaa	aatacagaaa	gaatattaat	ttaaaatttt	gaaagaggag	86400
tgatctgaac	ttatatctta	aaaagatcat	tctagggcat	ggtgggtcat	gcctgtaatc	86460
aagggttttg	ggaggctgag	acaggaggat	cacctgaggg	cagttcgaga	tcaacctgta	86520
cagcatagag	agactccatc	tctacaaaaa	gaaaaataaa	atagctgggt	gttgtgtggt	86580
attcaggagg	ctgaagcaga	aagatcactt	gagcccagga	gtttgagggt	gcagtaagct	86640
atgatcccac	cactgcaaca	cagtgaagtc	ttgtctcaaa	aaaaaaaaaa	aatcattcta	86700
ggtgcttttt	ggaggctgga	tgtggttaaga	gtagaagctg	gagatgggtc	tgttagggat	86760
tcgattcaga	ctttaaatc	catcaatgca	ttgagtccca	aatttacatc	actacgttgg	86820
atccttgccc	ctgaatccag	actggatat	ccaactttag	gttcagtttg	tatctctacc	86880
tgaccaatat	agagggtgtc	agtcttttgg	cttccctagg	ccacattgga	agaagaattg	86940
tcttgagcca	cacatagagt	acactaacgc	taacaatagc	agatgagcta	aaaaaaaaatc	87000
gcaaaaactta	taatgtttta	agaaagttta	cgaatttgtg	ttgggcacat	tcagagccat	87060
cctgggcccgc	gggatggaca	agcttaatcc	agtagatacc	ttcaacttac	aatatctaaa	87120
attttatgcc	agatttagtc	attttaaacc	tgctcatcag	tttttctcaa	gaagtagtat	87180
tttggttttt	tttcttttct	tttttttgag	atggagtttc	gctcttatcg	ttcaagctgg	87240
agtgcagtgg	cggatcttgg	ctcactgcaa	cctccgcctc	ctgggttcaa	gtgattctcc	87300
tgccctcagcc	ctgcaagtat	ctggaattac	aggcatgcgc	caccatgacc	agctaatttt	87360
tggagacagg	gtttccacct	gttgggtcagg	ctggttttgt	actcctgacc	tcaggtgac	87420
tgccctgcctc	ggcctcccaa	aggctgggat	tacaggcatg	agccaccgct	cccggctgca	87480
tttttgattt	tttagttgct	cagcccaaaa	cttttagtaca	tctttgaacc	tcttctttcc	87540
tccactacta	tatctgatcc	atcagcaaat	ctgttaggtc	tacctcacac	atatcgaaat	87600
cctaccacgt	ctcaccatct	gtgacaatta	acaccctgg	ctaggcagtc	atctctgtta	87660
agattgagtg	gttaaggatg	tcctctaagg	agatgacatt	caaactcttag	cttaaagtgc	87720
aagagggagc	tgggtttata	aagattgagg	aggcagcatt	attttgccat	aggcttccat	87780
ttgggttcca	ttccattctt	gatacttatg	gtatatattc	aaaacaaatg	cacagaaaca	87840
gaccaggtta	tattgggaat	ttcggatata	gagttccctag	ttgggaaaag	atagactgat	87900

86/122

ctgtaaatga	tgctagttat	ccatcatctg	gcaaaaaata	atttcctgcc	tccctctcata	87960
tatctcagat	caacagactt	tttctgttaa	gggccaaatc	ataaatatatt	taggctttcc	88020
agaccatag	gtttctgtca	cactctcctt	tatccttgaa	gccatagaca	atatgtaaac	88080
aaatgggcat	ggctgtgcta	cgataaaact	ttacttacaa	aaactggtag	tgggccagtt	88140
taggcattggc	cagcactttg	ggaggctaag	gcagatggat	cacttggggg	caggagtttg	88200
agaccagcct	ggccaacatg	gtgaaaccct	gtctctacta	aaaatacaaa	aaatagctgg	88260
gcatgggtgg	gggtgtctat	aattccagct	actctggagg	ctaagacaca	agaatcactt	88320
gaaccagga	ggcagaggtt	gcagtggact	gagatagcac	cactgcactc	cagccagggg	88380
gacggagtct	taaagcaaaa	caaaacaaaa	ggtagtgggt	tgtatttggc	ccatgggctg	88440
tagtttgcca	atccctgatg	cagaaacaaa	tccaggtaa	ataagagcct	ggaatgttaa	88500
aaaaacaaaa	cttgaagtca	tgtagaagaa	caggtagggg	gaacaatcct	gatctcagga	88560
taggaaggga	tattgtctaa	aataagacac	aggaaaatat	aatccatgtt	gtgtaaattt	88620
gactacgtta	aaacttaaaa	ctttcgccaa	gcgcggtggc	tcacgcctgt	aataccagta	88680
ctttgggagg	ccgaggtgag	cagatcacca	ggtcaggaga	ttgagaccat	cctggctaac	88740
acggtgaaac	cccgtctcta	ctaaaaatac	aaaacattag	ccgggcgtgg	tggcgggcgc	88800
ctgtagtccc	agctacttgg	gaggctgagg	caggagaatg	gcctgaaccc	gggaggcgaa	88860
gcttgacagt	agctgagatc	gcgccactgc	actccagcct	gggcgacaga	gtgagattcc	88920
gtctcaaaaa	aacaaaacaa	aacaaagcaa	aaaacctaaa	actttcatac	aataaagtat	88980
acctaagata	cttctagaag	agaagattta	catccaggac	gtgtatggaa	tttctgcaag	89040
taataagtaa	aagacaaggg	acatgaagag	gcagttcaca	aaagaggaag	ccaaaatgac	89100
caataaacat	gaaaggatgt	ttaacctcaa	aggaaacaag	gaaatgaatt	aaaaacatca	89160
aatgccattt	caaaaactagt	aagttggcaa	aattaaaaat	accaaggatg	agaatatgaa	89220
gcatggctat	atgagtgcac	ggaatggtag	agtcactttc	attaaaaatg	cacataattt	89280
gtttttttat	tatttttttg	agacagtcta	gtcgcgccag	gctagaatgc	agtggcatga	89340
tctcggtcca	ccacaatctc	tgccctcctg	gttcaagcaa	ttctcctgcc	tcagcctcct	89400
gagtagctgg	gattacaggg	acatgccaca	acgcccgggt	aagttttgta	tttttagtag	89460
agacaggggt	ttgccatgtt	ggccaggett	gtctcgaact	cctgacctca	ggtgagctgc	89520
ttcccaaaat	gctgggatta	gaggcgtgag	ccaatgctcc	tggctgaaaa	aaatgcacat	89580
aattttgttac	ctagcaattc	catgtctaga	ggcttatcct	agagaaaattc	ttgcttatat	89640
gcataggaag	acgtgtacta	gaatgttcac	tagttgaatg	tttaagtga	aattaggaaa	89700
taaagtaaat	gttcattaac	aggaaaatga	gtaaagggtat	atttataaaa	caattaaagta	89760
gctaaaaatga	ataaaactaga	gctgcgtgaa	tgaactagaa	ctgggtcaat	agtcatgtca	89820
gattatttgaa	tgaatcacagg	tcagatatgt	atagagtgtc	atttgtgtaa	ttaatttttt	89880
tttttttttt	gagatggagt	ctcactctgt	tgcccagggt	ggagtgcagt	ggcgtgatct	89940
cagctcactg	caacctccac	ctcctgggtt	aaagtgatcc	tcctgcctca	gcctcccagag	90000
tagttgggat	tacaggcatg	caccaccatg	cccagctcat	tttctatatt	ttagtggcca	90060
cagggtttca	ccatgttggc	caggctgggc	ttgaactcct	gacctcaagt	gttccaccca	90120
acttggcctc	ccaaagtgtc	aggattacag	gcgtgagcca	ccgtgctcag	ccatttgcgt	90180
gattttttaa	gatgtgcaga	ataatgccat	taaaaaaaat	acacatacat	gtatatatat	90240
acacgttttg	ctgggtgtgg	tggctcacac	ctgtaatccc	agcactttgg	gaggctgagg	90300
caggaggatc	acttgagccc	aggtgtacaa	gactagcctg	ggcgagatag	caagacccca	90360
tctcaacaac	agaaaaggata	attaggtatg	gtggcatgag	aggatcactt	gagcccagga	90420
gttcgagtgt	tatcaggcca	ctgcactcta	gcctggacaa	caaagcaaga	ccgtgtctca	90480
aaaaaataaa	aataaaaaagt	atttgtatgt	ggtcatagtc	aaaaaacgta	catggaagga	90540
aaatgtcttt	atttattttat	ttattttttt	ttttttaaga	cagagtcttg	ctctgtcacc	90600
caggctgggg	tacagtgggtg	taatctcagc	tcaccgcaat	ctcggcctcc	cgggttcaag	90660
cgattctttt	gcctcagcct	tctaagttagc	tgggactaca	ggtaccgcgc	accacaccct	90720
gctaattctt	gtgtttttcag	tagagacagg	gtttcaccat	gttggcaagg	ctggtctcga	90780
actcctgacc	ttaagtgagc	caccgcctt	ggcctcccaa	agtcctggga	ttacaggtgt	90840
gagccactgc	gcttggccag	gaaatatcta	atttagtaag	tatttatatc	tgggaaagga	90900
agggtcaggt	ggtagttcat	aggaaactcta	aagtctatgt	ataatactta	gggggacaga	90960
aggaaataaa	gcaaaatgct	gatatttgat	tggtgagttg	tgtatatgtt	agaagtataa	91020
cataggagat	ctgattgata	gtaggagaat	gttttttaggt	ggtaaaagtg	gaaccgtggg	91080
ggtttgtttt	ggcagtagaa	tcagttgggtc	atagtttgta	tgtggaagg	aataaacaga	91140
ccatgtttaag	gatgacttcc	ggaattttgg	tctgagttag	gggtggatga	cagtgtcatt	91200
catgaggga	gatgaagact	gaggtaggaa	caggtttggg	agaagatgac	atgttccctt	91260
ttagacaagt	ggaatttatgg	aagatggcag	gtaggtgggt	agctatatga	atttgagata	91320
aaagatttag	gatggagata	taaatttagg	agtaacagcg	tatctatggg	attgtaagcc	91380
ttaagaatgg	gtaggatcag	ccaggaaata	cagatgtata	tgcagaagag	aggagtcaag	91440
gaagccaaga	caagtttaatg	tttaagtga	gtgatgtagt	ccatgggcag	atgctgctga	91500

87/122

gagggctgca	aacaccagtg	accctacaac	atTTTTtaaat	gtcgtcttcc	tgacagcagt	91560
gatcagtagc	tgcaacgac	ttattttattt	TTTTcatgtt	agtctccaca	cacttgaatg	91620
tagacttttt	gaaggcaaaa	tcattgcctt	ttctgagctg	ggagcatgtc	tgccacatac	91680
caagcactca	acagttgatg	tattgacttc	atccagatac	tctgagggcg	agttatttcc	91740
tgctactagc	ctttcacctt	tcaatgttta	agagcacaaa	tacagagatg	ggcacgtttt	91800
ggcatttctt	atTTTTgataa	ccttttctctg	gtaagatttt	ttaatgttga	aaaaaaaaaa	91860
caagaaaaga	gggttaaaaa	tagtcttatg	tcagatcctg	tgatagaatt	cacacttggc	91920
ttaagctgct	gggcaccttc	ctatcttgga	tgctcatatta	gcttatctac	agcagaattt	91980
ttactgtttt	atgtagtaag	gaagcaatta	tatgattatt	ttacagacaa	attattcttt	92040
atctttttatt	tttttagacg	gagtctctct	ttgtctccca	ggctggagta	cagtgtcgcg	92100
atctcggtc	actgcaacct	ccgcctcctg	gggtcaagca	attctctgcc	tcagcctccc	92160
aagtagctgg	gcttacaggt	gtccgccacc	acaccagct	cattgttttg	tatttttagt	92220
agagatgggg	tttcaccatg	ttggccaggc	tggtcttgag	ctactgacct	caggtgatcc	92280
acccgccttg	gcatcccaaa	gtgctggaat	tacaggcgtg	agccaccgtg	cctggcccag	92340
acaaattatt	atactctgag	tgttagaggc	ttagatgtt	ttcacttgat	gctatgggag	92400
gaataagtaa	taagatatga	tacacaacca	aagacctttc	ttcactatgc	ttctagtagc	92460
tagtactatg	gatgacacat	ggtaataata	ttgggttagca	tttgtcctca	atttactgtg	92520
ctagttaact	ttctaagccc	cttacaggta	tatatttttt	ttcatcaata	atcctctaag	92580
tagtTTttta	ttattgacct	aattttataa	atcaagaaaa	ttaagacca	gagaagtaag	92640
taactttgtc	aagatcacat	ggcttataag	tggtagagcc	agaatttgac	cccagatgtt	92700
gtgactacat	tgtctctcca	taagcagggt	caactctttt	gactggatgc	tggtccaagg	92760
tcacttccct	agagaagcct	ttgctgacaa	ctaccctcct	gtgccctcct	ccaaggctgt	92820
ccattgttct	agaactttga	atactcatct	tagaataaag	ctggtcta	ttttacagt	92880
ttatagaatg	gatctctgac	tgcaaaagtt	ggtcataa	atctttttat	gttctagtga	92940
aaggcaaaaga	acaagagaag	acctcagatg	tgaagtccat	taaaggtaag	ttctgcctt	93000
ggcagtcac	tgcatataaa	agtgatgtgc	tttgcatTTg	tgagttcttt	aatcctgtta	93060
tactctctct	tttggcatta	atcatttctg	ccttattttta	taattactta	tgattttgat	93120
ttatttccct	ctttaacctg	tataatgctt	taacatctag	catataataa	gtaggctttt	93180
tttttttttt	tttttttggg	gacggagtct	tgctctgTTa	cccaggctgg	agtgcagtgg	93240
cgcgatcttg	gctcactgca	agctctgtct	ccggggttca	caccattctc	ctgcctcagc	93300
ctccccagca	gctgggacta	caggtgcacg	gcgccacgcc	tggctaattt	tttgtatttt	93360
ttagtagaga	cagagtttca	ccatgttagc	cagtatggtc	tcgatctcct	gaccttgtga	93420
ttcgcccccc	tcggcctccc	aaagtgctgg	gattacaagc	gtgagccacc	gcaccggcc	93480
gtaagtaggc	tttttttacc	ttaattttat	ttttttgaga	tggagtcttg	ctcttatccc	93540
caggctggag	tgcatgtgtg	ccatctcggc	tcactgcagc	atccacctcc	cgggttcaag	93600
cgattctctt	gcctcagcct	cccagtagc	tgggattaca	ggtggccgcc	accatgccc	93660
gctaattttt	gtatttttag	tagagacagg	gtttcaccgt	gttggccagg	ccagtctcaa	93720
actcctgacc	tcaagtgatc	cactgcctt	ggcctcccaa	agtcctggga	ttacaggcgt	93780
gagccaccat	gcctggccat	aagttaggctt	ttactgagcc	ttgtgtgtat	tggtctcct	93840
agtgattaca	gtgaaccagt	gcccttctta	ttaatcacac	atttaattgt	tccttaaaag	93900
tgattagttc	actttatttta	tttagtaaga	caaaaaatga	agaatactct	taactgagca	93960
gtctgttaac	tgtaggaaag	cactgacact	tataaggctt	agttttctgt	catttatcca	94020
gaagtatgg	tgattacagt	ttttactttt	ttatttgaat	gaacaacctt	aattttaa	94080
atattttgtt	tattttttgt	tgggatcgat	acattgtcct	tgtttataga	ttagagcatg	94140
cttttttaaa	atgctgtatt	actcactgat	tttatttgtc	cagtgtacag	agattgaagt	94200
gggaaaaata	taatggaaat	tgttccata	gtcattacat	attaatttca	tcaatttatt	94260
tcataaaaat	ctgtagattg	ctacttattt	agatttttcc	ttcaaattgt	tttatgttgt	94320
attgcttgca	ctgagtattt	attctatatg	ctcaatttgc	tggagaagaa	gactaattat	94380
aacttaggca	agttgtaaaa	ttagggaaaa	aagtaaggta	ccttacagcc	tagtttactt	94440
atttcttatg	taaagccagt	tagattccac	attagtcaaa	actgccttct	ttgagcaaaa	94500
cttgattggc	agtgataaa	gcttaaagcc	cttctcaagc	agagacctgt	aaagactaga	94560
tctgactgta	gtagaaggaa	ggaacttaga	tgtttcaggc	agtgagaaca	ccagtcttcc	94620
actctaaact	ttgccactaa	cagtatgacc	ttgggaagtt	gtaactttct	tcagattctt	94680
catttggtga	atggggggat	tggcctagct	aatttctaaa	tctctactgg	gctaaaaaat	94740
tctgtgctta	tactctgatt	atgaagtaca	taatctgtgc	ttaacattca	ctgacttatc	94800
cttaggataa	tacagaagca	gtacaagaaa	cagccctcca	agatgtttgc	agctctgtta	94860
gaaagacaaa	cttatacaca	gaacagtagc	aaatagacca	aaataataat	agctgccatt	94920
tatagaacac	ttcttctgtt	ctgggcatta	gacaaaaact	gactataacg	gtgaacaaaa	94980
aagacttagg	tcctgccctc	attgaactta	cagattagta	ggggagagga	acattaatca	95040
agtaattcca	cagatggcct	agcctagatt	ggtagtgtg	gaagtaaaga	gatgtgaacg	95100

88/122

gacttgaaaa	aaaatttcgga	ggcaaaatgg	atagaagttt	attattgatt	aaatatgagg	951
tgtgagagag	agggatat	aagattgata	cctaccttct	ggcttgccct	acagaaccaa	952
aacaggaaat	tatatgttca	gttttgttat	gttgggtggg	aggtgctttt	gagtcattca	952
tttatatatg	ttatatatgt	tattttatat	gcatagtaat	tttaaggtct	gagttttaa	953
ccaaagggtta	gagagtgtat	tttttagagtc	tagcaaacct	aagttgaaat	cctgcctgtt	954
gaaatggctg	tttactagct	cattaacctta	gggcaaagta	ttcaacttgt	tttcatTTTT	954
gtcttcatct	ctaaaatgag	gaaaatatgg	tcttacaaga	ttgtcctgag	agatagatga	955
aataatatcc	aaaaaaaaaa	aagggtacata	gagaaactcg	tatagtgcct	ggtatatagt	955
aggctctcca	ttggtagcta	tcattatcta	gttttaacat	agccttcagt	ttgttgaatt	956
agtcaaaactg	agtgaagcac	tgcaagggaat	tcagagggaat	ttgagatcaa	caaattgattt	957
ctgaagttaa	gggaagactt	catggcaatg	acactttacct	tgtataaaaag	ttgaagaata	957
agaaagattt	agtagagaga	ttctttctct	tctccctacc	agcccagctt	cttatttgag	958
gatatttggt	gcaaaggggc	cttcagacaa	gtagagggag	attttttacag	aaagattgag	958
atgaaggat	agaaggctgt	aaagaccaga	aaagagaatt	gagacagagg	aagcagggaag	959
ccactgtagg	tttttgagca	agatattgat	gctgtaagta	tggtgtttat	gaaaggttag	960
tctggaagag	atttgcagga	tggagacccc	gtgcttgaga	ttgttataat	acagaaagac	960
rrgactgag	ggtagggtgt	taaaaataaa	caggtaagta	aatgttttaa	catcttgaag	961
qaaaaagtc	caaatcttgg	caagtaaaaca	gataacagtg	aaaaagaatg	ggaccaagat	961
tttgagtttt	ggagactggg	ggattgaaca	gacagggaaa	ttgagaggag	aatcagatga	962
tgatgtttta	agttgatatt	tagacagatt	gtgcttgaga	tggtaaaagtc	aatgtgggtg	963
ggaatgctta	gtagcgagta	atcagtata	caagaccaaa	gcccagggtca	aagacaagtc	963
acagatacag	atcagggctt	tttcatctgc	tccacagagg	tgtaccctag	gagctgttgc	964
aaacagctcca	tgtggagggt	gtgagtaaga	tgtttccctt	gaatttgcca	gaattacttt	964
tttgtgttg	ttgttgtttt	ttctgagaca	gttctgcgt	ctgttgccca	ggctggaggg	965
cagtggcgag	atcgcgagc	tactgcaac	ctctgcctct	cgggttcgag	tgattctcct	966
gcctcagcct	cccaagtagc	tgggattaca	ggcttgtgcc	accaagccca	gctaatttct	966
tttgtatttt	tagtagagat	gggggtttcac	catgttggcc	agactgggtct	cgaactcctg	967
gcctcgtgat	ctgcctgcct	cagcctccaa	aagttctggg	attacaggcg	tgaaccactg	967
cacccggctcc	cttgttaagt	ttattttggg	gggaagcaaa	ggaggtttca	gcttttaaaa	968
agtttgaaaa	ttattgtctt	ggtaaaat	aaagatttga	gagtaaatat	gctttctagc	969
agaaagaata	aaagaagaac	agatagcctc	aagaagggga	gccaaagaag	caggctatat	969
ctgacacact	gggtgttgat	aaatgggtat	taaaagaatg	agagcaatga	gcagatagaa	970
gaggaataa	ggagagtata	ataccatgga	gaccaagaaa	gatagactat	caggaaggag	970
tggtaaaaat	aagttaactag	ttctaagaga	tatgttaaga	gggaccgggg	aaagccttgt	971
acaaatgagt	tagtagcatt	ttacattata	gatctaat	taagaaacaa	tgcgagagtc	972
tcaccattcc	tactgtact	tactgtact	tgtctgaaca	cgaaaactgg	cttttgttta	972
taaataagct	aaaaattatt	ttgctccaat	ttctcatgaa	ataaaaaata	aaccttcttt	973
taacattgaa	aaaatagttt	gaagacagtc	actcttcatt	ttgtaattcc	cacaactatt	973
attgaatgac	tgaattatc	tttattctga	agccaaaggg	gtgatactga	tattttctca	974
gactactaaa	aatatatttt	atgaattttt	agtggtgctt	atcttttttt	gttttttttt	975
ttgagatgga	gtttcactcc	cgttgctcag	gctggagggg	agtggtgcaa	tctcagctca	975
ctgcaacctt	cgctccccag	attcaagcaa	ttctcctgcc	tcgggtctccc	aagttagctgg	976
gattacagge	acctgcccc	acacccagct	aattttttgt	atttttagta	gagacagggt	976
tlcaccatgt	tggtcaggct	ggctctgaac	tcctgacctc	aggtgatcca	cccaccttgg	977
cctcccaaag	tactgcgatt	gcaggcatga	gccaccatgc	ctggcctgag	gaatattttt	978
ctaggttccc	cccaccccaa	gcatttatct	tgcaatttta	gttttgttcc	taaagcaagc	978
aaggttttaag	gatttataaaa	taatccgtat	tttagaatgc	tttctggctt	tgttactttt	979
tatcccaagt	agaagttctc	agagaatgat	ctccctcttt	taatttaact	ttttggcaca	979
gtattttgag	aattataaat	aatattagaa	tgttttctgg	ctgggtgtgg	tggtcatgac	980
ctgtaatcct	ggctacttgg	gaggctgagg	cagggaatc	acttgaacat	gggaggcaga	981
gggttcagtg	agccgaggtc	atgccactgc	actccagcct	gggtgacaga	gccagactct	981
gtctgggaaa	aaaaaaaaaa	aaaaaaagag	tggtttcttt	cctattttcc	accacttgat	982
taagttactt	ttcctcttaa	gtattttttg	ctgagtatgc	tgacttaaga	gtaattgtac	982
aaaatttaatt	ttttaagttt	ctctgaaagc	ccctttatga	gagtttttagg	ctatcaaat	982
gtgttttaatt	cttaacaatt	ttttgaaaaa	tatatgcttc	aatatccgta	cattccccac	984
aaaaaagcac	taaaaatcat	gccttgctgg	aggctgcagg	accaagtcac	gttgcaatca	984
atgccatttc	tgccaacatg	gactcctttt	caagtagcag	gacagccaca	cttaagaagc	985
agccaagcca	catggaggcc	gctcattttg	gtgacctggg	taagtaacta	tcatttttta	985
ttaaactgta	ttagaaggat	ttgagtacaa	tatgtgaaac	ttctgtcata	ggatacagaa	986
ctatataaatt	ggaaagtgtc	ttggaaaaaa	tgtattttaa	ataacagcta	caagtataat	987

87/122

gagggctgca	aacaccagtg	accctacaac	atTTTTaaat	gtcgtcttcc	tgacagcagt	91560
gatcagtacc	tgcaacgac	ttatttattt	ttttcatgtt	agtctccaca	cacttgaatg	91620
tagacttttt	gaaggcaaaa	tcattgcctt	ttctgagctg	ggagcatgtc	tggcacatac	91680
caagcactca	acagttgatg	tattgacttc	atccagatac	tctgagggcg	agttatttcc	91740
tgctactagc	ctttcacctt	tcaatgttta	agagcacaaa	tacagagatg	ggcacgtttt	91800
ggcattttctt	atTTTTgataa	ccttttctctg	gtaagatttt	ttaatgttga	aaaaaaaaaa	91860
caagaaaaga	gggttaaaaa	tagtcttatg	tcagatcctg	tgatagaatt	cacacttggc	91920
ttaagctgct	gggcaccttc	ctatcttgga	tgtcatatta	gcttatctac	agcagaattt	91980
ttactgtttt	atgtagtaag	gaagcaatta	tatgattatt	ttacagacaa	attattcttt	92040
atcttttatt	tttttagacg	gagtctctct	ttgtctccca	ggctggagta	cagtgtcgcg	92100
atctcggtct	actgcaacct	ccgctcctctg	ggttcaagca	attctctgcc	tcagcctccc	92160
aagtagctgg	gcttacaggt	gtccgccacc	acacccagct	cattgttttg	tatttttagt	92220
agagatgggg	tttcaccattg	ttggccaggc	tggctctgag	ctactgacct	caggtgatcc	92280
acccgccttg	gcatcccaaa	gtgctggaat	tacaggcgtg	agccaccgtg	cctggcccgag	92340
acaaattatt	atactctgag	tggttagaggc	ttaggatgtt	ttcacttgat	gctatggggag	92400
gaataagtaa	taagatatga	tacacaacca	aagacctttc	ttcactatgc	ttctagtagc	92460
tagtactatg	gatgacacat	ggtaataata	ttggttagca	tttgtcctca	atttactgtg	92520
ctagtacttc	tttttagccc	cttacaggta	tatatTTTTt	ttcatcaata	atcctctaag	92580
gtagttttta	ttattgacct	aattttataa	atcaagaaaa	ttaaagacca	gagaagtaag	92640
taacttgctc	aagatcacat	ggcttataag	tggttagagcc	agaatttgac	cccagatgtt	92700
gtgactacat	tgtctctcca	taagcagggt	caactctttt	gactggatgc	tgttccaagg	92760
tcacttcctt	agagaagcct	ttgctgacaa	ctacccctcct	gtgccctcct	ccaaggctgt	92820
ccattgttct	agaactttga	atactcatct	tagaataaag	ctggctcta	ttttacagt	92880
ttatagaatg	gatctctgac	tgcaaaagtt	ggctcataat	atctttttat	gttctagtga	92940
aaggcaaaaga	acaagagaag	acctcagatg	tgaagtccat	taaaggtaag	ttctgccctt	93000
ggcagtcac	tgcattaaaa	agtgatgtgc	tttgcatTTg	tgagtTcttt	aatcctgtta	93060
tactctctct	tttggcatta	atcatttctg	ccttattTTa	taattactta	tgattttgat	93120
ttatttccct	ctttaacctg	tataatgctt	taacatctag	catataataa	gtaggctttt	93180
tttttttttt	tttttttgga	gacggagtct	tgctctgtta	cccagggtgg	agtgcagtgg	93240
cgcgatcttg	gctcactgca	agctctgtct	cccgggttca	caccattctc	ctgcctcagc	93300
ctccccagca	gctgggacta	caggtgcacg	gcgccacgcc	tggctaattt	tttgtatttt	93360
ttagtagaga	cagagtttca	ccatgttagc	cagtatggtc	tcgatctcct	gaccttgtga	93420
tccgcccggc	tccggcctccc	aaagtgtctg	gattacaagc	gtgagccacc	gcacccggcc	93480
gtaagtaggc	tttttttacc	ttaatTTtat	ttttttgaga	tggagtcttg	ctcttatccc	93540
caggctggag	tgcatgtgtg	ccatctcggc	tcactgcagc	atccacctcc	cgggttcaag	93600
cgattctcct	gcctcagcct	cccagtagc	tgggattaca	ggtggccgcc	accatgccca	93660
gctaattttt	gtatttttag	tagagacagg	gtttcacctg	gttggccagg	ccagtctcaa	93720
actcctgacc	tcaagtgatc	cactcgcctt	ggcctcccaa	agtcctggga	ttacaggcgt	93780
gagccaccat	gcctggccat	aagtaggctt	ttactgagcc	ttgtgtgtat	tggctatcct	93840
agtgattaca	gtgaaccagt	gcccttctta	ttaatcacac	atttaattgt	tccttaaaag	93900
tgattagttc	actttatttta	tttagtaaga	caaaaaatga	agaatactct	taactgagca	93960
gtctgttaac	tgtaggaaag	cactgacact	tataaggctt	agttttctgt	catttatcca	94020
gaagtatggt	tgattacagt	ttttactttt	ttatttgaat	gaacaacctt	aatttaaaat	94080
atattttgtt	tattttttgt	tgggatcgat	acattgtcct	tgtttataga	ttagagcatg	94140
ctttttaaag	atgctgtatt	actcactgat	tttatttgtc	cagtgtacag	agattgaagt	94200
gggaaaaatta	taatggaaat	tgtttcata	gtcattacat	attaatttca	tcaatttatt	94260
tccataaaaat	ctgtagattg	ctacttattt	agatttttcc	ttcaaagtgt	tttatgttgt	94320
attgcttgca	ctgagtattt	attctatatg	ctcaatttgc	tggagaagaa	gactaattat	94380
aacttaggca	agttgtaaaa	ttagggaata	aagtaaggta	ccttacagcc	tagtttactt	94440
atttcttatg	taaagccagt	tagattccac	attagttcaa	actgccttct	ttgagcaaaa	94500
cttgattggc	agtgataaag	gcttaaagcc	cttctcaagc	agagacctgt	aaagactaga	94560
tctgactgta	gtagaaggaa	ggaacttaga	tgtttcaggc	agtgagaaca	ccagtcttcc	94620
actctaaact	ttgccactaa	cagtatgacc	ttgggaagtt	gtaactttct	tcagattctt	94680
catttggtga	atggggggat	tggcctagct	aatttctaaa	tctctactgg	gctaaaaaat	94740
tctgtgctta	tactctgatt	atgaagtaca	tactctgtgc	ttaacattca	ctgacttatc	94800
cttaggataa	tacagaagca	gtacaagaaa	cagcccttca	agatgtttgc	agtctggtta	94860
gaaagacaaa	cttatacaca	gaacagtagc	aaatagacca	aaataataat	agctgccatt	94920
tatagaacac	ttcttctgtt	ctgggcatta	gacaaaaact	gactataacg	gtgaacaaaa	94980
aagacttagg	tcttgccttc	attgaactta	cagatttaga	ggggagagga	acattaatca	95040
agtaattcca	cagatggctt	agcctagatt	ggtagtgtg	gaagtaaaga	gatgtgaacg	95100

88/122

gacttgaaaa	aaaattcggga	ggcaaaatgg	atagaagttt	attatttgatt	aaatatgagg	95160
tgtgagagag	aggggatattt	aagattgata	cctaccttct	ggcttgccta	acagaaccaa	95220
aacaggaaat	tatatgttca	gttttgttat	gttgggtggg	aggtgctttt	gagtcattca	95280
tttatatatg	tatatatgt	tattttatat	gcatagtaat	tttaaggtct	gagttttaaa	95340
ccaaagggtta	gagagtgtt	tttttagagtc	tagcaaacct	aagttgaaat	cctgcctggt	95400
gaaatggctg	tttactagct	cattaaccta	gggcaaagta	ttcaacttgt	tttcattttt	95460
gtcttcatct	ctaaaatgag	gaaaatatgg	tcttacaaga	ttgtcctgag	agatagatga	95520
aataatatcc	aaaaaaaaaa	aaggtacata	gagaaactcg	tatagtgcct	ggtatatagt	95580
aggtcctcca	ttggtagcta	tcattatcta	gttttaacat	agccttcagt	ttgttgaatt	95640
agtcaaactg	agtgaagcac	tgcaaggaat	tcagaggaat	ttgagatcaa	caaatgattt	95700
ctgaagttaa	gggaagactt	catggcaatg	acacttacct	tgtataaaaag	ttgaagaata	95760
agaaagattt	gaatgagaga	ttctttctct	tctccctacc	agcccagctt	cttatttgag	95820
gatataattg	gcaaaggggc	cttcagacaa	gtagagggag	attttttacag	aaagatttgag	95880
atgaagggtat	agaaggctgt	aaagaccaga	aaagagaatt	gagacagagg	aagcaggaag	95940
ccactgtagg	tttttgagca	agatattgat	gctgtaagta	tgggtgtttat	gaaagggttag	96000
ctcggaagag	atttgcagga	tggagacccc	ggaagttttt	ttgttataat	acagaaagac	96060
ttgcactgag	gggtgaggtgt	taaaaataaa	caggtaacat	aatgtttaaa	catcttgaag	96120
gaaaagtcaa	caaactcttg	caagtaaaaca	gataacagtg	aaaaagaatg	ggaccaagat	96180
tttgagtttt	ggagactggg	ggattgaaca	gacaggggaa	ttgagaggag	aatcagatga	96240
tgtgttttta	agttgatatt	tagacagatt	gtgcttgaga	tggtaaagtc	aatgtgggtg	96300
ggaatgctta	gtagtcagta	atcagtgata	caagaccaaa	gccagggtca	aagacaagtc	96360
acagatacag	atcagggtct	tttcatctgc	tccacagagg	tgtaacctag	gagctgttgc	96420
aaacagtcca	tgtggagggg	gtgagtaaga	tgtttccctt	gaatttgcca	gaattacttt	96480
tttgttgtg	ttgttgtttt	ttctgagaca	gattctctgt	ctgttgccca	ggctggaggg	96540
cagtggcgag	atcgcgagc	tcactgcaac	ctctgctct	cgggttcgag	tgattctcct	96600
gcctcagcct	cccaagtagc	tgggattaca	ggcttgtgcc	accaagccca	gctaatttct	96660
tttgtatttt	tagtagagat	gggggtttcac	catgttggcc	agactgggtct	cgaactcctg	96720
gcctcgtgat	ctgcctgcct	cagcctccaa	aagttctggg	attacaggcg	tgaaccactg	96780
caccgggtcc	cttgttaagt	ttatttttgg	gggaagcaaa	ggaggtttca	gcttttaaaa	96840
agtttgaaaa	ttattgctct	ggtaataatt	aaagatttga	gagtaaatat	gctttctagc	96900
agaaagaata	aaagaagaac	agatagcctc	aagaagggga	gccaaagaag	caggctatat	96960
ctgacacact	gggtgttgat	aaatgggtat	taaaagaatg	agagcaatga	gcagatagaa	97020
gaggaaatta	ggagagtata	ataccatgga	gaccaagaaa	gatagactat	cagggaaggag	97080
tggtaaaaa	aagttactag	ttctaagaga	gatgttaaga	gggaccgggg	aaagccttgt	97140
acaaatgagt	tagtagcatt	ttacattata	tacatcta	taagaaacaa	tgcgagagtc	97200
tcaccattcc	tatagactct	tacttgtact	tgtctgaaca	cgaaaactgg	cttttgttta	97260
taaataagct	aaaaattatt	ttgtccaat	ttctcatgaa	aataaaaaata	aaccttcttt	97320
taacattgaa	aaaatagttt	gaagacagtc	actcttcatt	ttgtaatttc	cacaactatt	97380
attgaatgac	tgaattatc	tttattctga	agccaaaggg	gtgatactga	tatttcttca	97440
gactactaaa	aatatatattt	atgaattttt	agtgtgcttt	atcttttttt	gttttttttt	97500
ttgagatgga	gtttcactcc	cgttgctcag	gctggagggc	agtggtgcaa	tctcagctca	97560
ctgcaacctt	cgctcccag	attcaagcaa	ttctcctgcc	tcggtctccc	aagtagctgg	97620
gattacaggc	acctgcccc	acaccagct	aattttttgt	attttttagta	gagacagggt	97680
ttcaccatgt	tggtcaggct	ggtcttgaac	tcttgacctc	aggtgatcca	cccaccttgg	97740
cctcccaaag	tactgcgatt	gcaggcatga	gccaccatgc	ctggcctgag	gaatattttt	97800
ctagggtccc	cccaccccaa	gcatttatct	tgcaatttta	gttttgttcc	ttaaagcaagc	97860
aagggtttaag	gatttaaaaa	taatccgtat	tttagaatgc	ttctgggtct	tggtactttt	97920
tatccacagt	agaagtctc	agagaatgat	ctccctcttt	taatttaact	ttttggcaca	97980
gtattttgag	aattataaat	aatattagaa	tgttttctgg	ctgggtgtgg	tggctcatgc	98040
ctgtaatcct	ggctacttgg	gaggctgagg	caggagaatc	acttgaacat	gggaggcaga	98100
ggttgcagtg	agccgaggtc	atgccactgc	actccagcct	gggtgacaga	gcaagactct	98160
gtctgggaaa	aaaaaaaaaa	aaaaaaagag	tgttttcttt	cctattttcc	accacttgat	98220
taagttactt	ttctctttaa	gtattttttg	ctgagtatgc	tgacttaaga	gtaatgttac	98280
aaaatttaatt	ttttaaagtt	ctctgaaagc	ccctttatga	gagttttagg	ctatcaaatt	98340
gtgtttaatt	cttaacaatt	ttttgaaaaa	tctatgcttc	aatatccgta	cattccccac	98400
aaaaaagcac	taaaaatcat	gccttgctgg	aggctgcagg	accaagtcac	gttgcaatca	98460
atgccatttc	tgccaacatg	gactcctttt	caagtagcag	gacagccaca	cttaagaagc	98520
agccaagcca	catggaggcc	gctcattttg	gtgacctggg	taagtaacta	tcatttttta	98580
ttactttgta	ttagaaggat	ttgagtacaa	tatgtgaaac	ttctgtcata	ggatacagaa	98640
ctatataatt	ggaaagtgtc	ttggaaaaaa	tgtatttttaa	ataacagcta	caagtataat	98700

89/122

gggtagctgt	gttgtgttcc	tgtaaatata	gaatataaa	catgcccagt	agaaaaacaa	98760
gcatttccag	aagaaatata	tctgatcact	aaatataaat	atatgaaaaa	gatgtctcac	98820
tttattactg	agggaaagtgc	aaattaaaa	aatcagttaa	tgttctccta	acacattagc	98880
atatttttta	aagtttgaca	atgtgaatgt	cagtgaagat	gcagggaat	accctccta	98940
tttagtgata	atataatctg	gtgaagactc	tttggaagc	aatttgga	tcagtataaa	99000
atatgcatgt	catttaggcc	actctttcta	agacctagcc	ctcagatatg	ctcattcata	99060
tgtgcaggtg	tgtatgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtatatgta	tgtatgtatg	99120
tatgtatgta	tgtatgttga	aggctattca	ttatagtatt	gtttgtgata	gcaaaaaatt	99180
atggacaaca	tataaatatc	tgttataggg	aaataaccaa	attgtggtat	acgcatgtc	99240
tggagtataa	tatagccatt	tgtttctatt	tatttatatt	cttgagacag	ggttttactc	99300
tgttgcccag	gctggagtgc	agtggatga	tcatggttca	ctgcagcctt	cacctcctgg	99360
gcacaagcca	ttctctcgcc	tcagcctcca	gactgcaggc	gcactatggt	atgtgtcacc	99420
acaccagat	aattttttta	ttttttgtag	agacagggtc	tcactatggt	gcctaagctg	99480
gtctcaaaact	cctggcctca	agcaattctc	ccacacaggc	ctcccaaagt	gctgggatta	99540
ccaacgtgaa	ccaccacacc	tgggtcagtg	tagccattta	gaaatctaaa	aaagacgtgg	99600
gaaaaatgtct	aaggcatgtt	taaatgtgag	aaaagcaagt	cacagtatgc	atggtaaaa	99660
ccgttatatt	aaaataagtt	cttccaaaac	aaaaacatat	gcaggagacc	tttattttgt	99720
cagtatttct	tacccaaatt	tctgcactta	gaaaattgca	tgtcatgttg	tcataagttg	99780
aaaaaaaagat	ccatgaacca	atggacttct	aataaaatca	gtcctgcttt	tgacatctct	99840
ctctactttt	gtgtatattc	aaaccagagt	gtcaatgtgt	ttgtggggca	cacttagcaa	99900
taatacatag	cagacaaaat	gcatatagct	cagagagtaa	aattgtaagt	tttgtagat	99960
cactcataaaa	ttgctgatga	gaatttaaaa	tgggtgcagat	gctctggaaa	acaggcagtt	100020
tctttctttt	tttttttttt	tctttttgag	acagggtctc	actctgttgc	gcaggctgga	100080
gtacagtggc	gtgattacaa	ctcactgcag	cctcaccctc	ctcagggtca	ggtgatctct	100140
cctcagtcct	ctgagtagct	gggactatag	gggtgcacca	ccacgcctgg	ctaatttttg	100200
tatttttttt	tttttttttt	gtagagacgg	gggttcgcca	tggttcccag	gctgggtctca	100260
aactcctgga	atcaagcgat	ccacttgcgt	aggcctccca	aagtgtctgg	attacgggcg	100320
tgagctactg	tgcctggcct	aggcagtttg	tttgtttgtt	tggttgtttg	tttattttatt	100380
tgtagacgga	gtctcacagg	ctggagtgca	tggtcccaat	ttttggctca	ctgaacctc	100440
cgctcccag	gttcaagcta	ttctcctgcc	tcagcctcct	gagtagctgg	gatgacaggt	100500
gcctgccata	atgcctggct	gattttttgt	tatttagtag	atatgggggt	tcaccatgtt	100560
ggtcaggctg	gttttgaact	cctgacctca	ggtgatcagc	ccgcctcggc	ctcccaaagt	100620
gctgggatta	caggcatgag	ccgtcatccc	tggtgtgtgg	tttcttatga	cgtgaaacat	100680
gcaattacca	tatgacctag	cagttgcact	ctgtatttat	cccagataaa	tgaaaactta	100740
ccttccaata	aaaacctgtg	cacaaatgtt	catagcagct	taatattgaa	aaactggatg	100800
ttcttcagca	ggtgaatgaa	ctgggtcatt	cataccatgg	aataccattc	agcaataaaa	100860
aggaacaaac	tttaaccata	tttaaccacc	ggtgaacaa	tcaagggaat	tatgctgtca	100920
gacaaaaacc	agtccctaaa	gactacatat	agtatgattc	cgtttgata	atattcttga	100980
aatagagaaa	ttaagagaaa	tgaaaagatt	agtgtttgcc	agatgttaga	gacagggagg	101040
tgagaggggt	aagtgggtgt	agttataaaa	gtgcaacatg	agggatcttt	gtgatgttga	101100
agttgtatct	tggcagtgga	tcagaaatc	tcaatgtgat	aaaattacaa	agaactaaaa	101160
acaagaatga	gtatagataa	aactggggaa	atctgaacaa	gtagagtgt	tgtatcactg	101220
tcagtatctt	agagtgatat	tgtactatag	ctttgcaaga	tggtaccatg	ggagaaacta	101280
aagtgtacaa	gggatctcta	ggtattatta	tttttttaga	gatgggggtt	cactatgttc	101340
cccaggccgg	tcttgaactc	ctgggctcta	gtgatccgcc	tgccccagcc	tcctaaagta	101400
ctggaattac	aggcgtgagc	gacctgcct	ggccttttca	gtattgtatc	ttagaacttc	101460
atgtgaatct	agcattatct	catagaattt	aattaaaaga	aattgtaaac	ctcacagaag	101520
atcagaatct	cctcaagttt	gtgatgttga	caaagatgaa	ctagtgtgaca	ctgacagtaa	101580
gactgaggat	gaagacacga	cgtgcttcaa	aaaaatgatt	tgaatatcaa	tggattaaga	101640
agaactcttt	tgacaaattg	atgaaaccct	cagtcagttt	tataagaatg	cccatcttta	101700
tgatcatgct	atgaaagcca	attttttaaa	aaattttttg	tctttcctaa	caattagctt	101760
gtgggtataa	tttaaattta	gttaaatata	agataaatga	ttttttatta	agtttagttt	101820
cattttttcaa	ggtacgatct	caaagctact	ctttaacccta	ctatgaatga	ataatgctga	101880
gttcataaca	tctttgtaga	tatatccaca	attttccctc	aggataagtg	cctacaagtg	101940
gaattactgg	actgaaaata	atgcagtttg	ctaagacttt	gctatctgtt	cctgaatgct	102000
cctccaaaaa	ggttttgcca	gtttacatcc	tcatgaccag	cgaatgagag	tggtgcctat	102060
tttctgtg	cctgtttact	gcttaataat	ttttgaaaaa	aatctaattt	gacagacaaa	102120
aatgcattgt	atgttaattt	gcttttctgg	gaggttgagt	gaggttgagt	atagttttta	102180
atatttttat	tggccccctt	ggaactagta	tcataagttt	tttttcttaa	gaatttatgt	102240
agtctgggct	gggcgcagtg	gctcacgcct	gcaatcccag	cactttggga	ggccgaggtg	102300

90/122

ggtggattgc	cgaagggtcag	gagtttgaga	ccatcctgac	caacatgggtg	aaaccgaatc	102360
tctactaaaa	gtacaaaaaac	tagctcagcg	tgggtggcggg	tgcctgtaat	cccagctact	102420
taggaggctg	agtcaagaga	atcgcttgaa	cccggggaggt	ggagggttgg	tgcattgagc	102480
cgagatcgcg	ccattgctct	ccagcctagg	caacaagagt	gaaaagtcctc	aaaaaaaaaa	102540
aaaaaaaaaa	aaaaaagaat	ttacatgggtc	tgaattgccca	ttaaaagaga	tatgagaatt	102600
attgagtaac	aaataacttt	ttaataattt	aggcaagttt	tggacgattg	tactttgttt	102660
agaaaccaaa	agcatagtat	ttgtagtttt	tttattttact	ttagttgcta	ggaagtaaac	102720
tttattcaag	gtctctggta	ccagttgttg	ctaaaagtga	ttgactaatc	tgtcaatctg	102780
aaattatttg	ttgtctgaact	gctaattctt	ttgcttctat	cttttaggca	gatcttgtct	102840
ggactaccag	actcaagaga	ccaaatcaag	cctttctaag	acccttgaac	aagtcttgca	102900
cgacactatt	gtcctccctt	acttcattca	attcatggaa	cttcggcgaa	tggagcattt	102960
ggtgaaattt	tgggttagagg	ctgaaagttt	tcatttcaaca	acttgggtcgc	gaataagagc	103020
acacagtcta	aacacagtga	agcagagctc	actggctgag	cctgtctctc	catctaaaaa	103080
gcatgaaact	acagcgtctt	ttttaactga	ttctcttgat	aagagattgg	aggattctgg	103140
ctcagcacag	ttgtttatga	ctcattcaga	aggaattgac	ctgaataata	gaactaacag	103200
cactcagaat	cacttgctgc	tttcccagga	atgtgacagt	gccccattctc	tccgtcttga	103260
aatggccaga	gcaggaaact	accaagtttc	caggaaacc	caagaatctt	cctctacact	103320
tacagtagcc	agtagaaata	gtcccgcctt	tccactaaaa	gaattgtcag	gaaaactaat	103380
gaaaagttag	tatgtgattt	tcttgtgtgt	acatatgtgt	ctcactttct	ttttttaatt	103440
tactaagcag	aacttcagat	gaggaataaa	atgattggaa	tatttttttt	ctcctctaac	103500
tacttgtaaa	tttgggagaa	tttggagagt	gtagtagagt	cagatcagtg	tatggaaaag	103560
gagcaggagt	gactggacct	tctaagaagt	gtgttatcag	aattagtaaa	tgaagggtca	103620
aatgtcctac	ttttcccttc	cactgatttt	gacatcaaac	cattatccac	atagccttat	103680
ttcctccctc	ggtcttaatt	ttattaatat	tttactgcac	tttgagata	aaatttttaa	103740
aaaattttta	aaaattgcca	ataagtga	tttattaa	tcagtgttta	gtgtatat	103800
ggattttatt	tatttagtcac	aagacctttg	tgcaggtagt	aggcatgatt	atcttttttt	103860
ttttgagatg	gagtcttgct	ctgtcgccca	ggctggagtg	caatggcgcg	gtctcggtct	103920
actgcaacct	ccgggttcat	gccattctcc	tgcctcagcc	tcccaaata	ctgggactac	103980
aggcgcctgc	caccacaccc	ggctaatttt	tttgtatttt	tagtagagac	gggggttcac	104040
catgttcgcc	aggatgggtc	cgatctctcg	actttgtgat	ccgcctgcct	cggcctccca	104100
aagtgtctgg	attacaggca	tgagccaccg	cgcccggact	gattatctta	tttacacatg	104160
agaaaaccag	ggcttagaaa	ggttaggtaa	cttcccttag	gttgtacagt	aaatgtggac	104220
ctagaagcat	tttgacaaga	gcacctgttt	ttttttcttc	tctattagtt	tagaaattat	104280
atactcttaa	ttatcacctg	ggattttgat	tagcagcct	tcattgttct	tttcatctta	104340
aatgttcttt	gtgtcttaaa	gggctaagt	atcttctcag	atcttttagt	tactcatttc	104400
tcagtgaact	aaaatgaggt	ctaactctgt	actgaatcaa	gttttcagca	tgttatttcc	104460
ttcctccctc	cctccctcct	tccttccctc	aaccaggctc	ccgaggagct	gggattacag	104520
gcgcccgcga	ccactccttg	ctaattttta	tatttttagt	gagacggggt	ttcaccatgt	104580
tggtcaggct	gatcttgaac	tcctgacctc	aagtgaacca	cctgcctcgg	cctcccaaag	104640
tgctgggatt	acaggcatga	atcaccacac	ctgacggcat	gttattttca	tcgcaaagtt	104700
actgtaagct	gggagaagtg	gcacacactt	gtactcccag	ctactcagga	agcttaaggt	104760
gagaagattg	cttgagccca	ggagttttga	gaccaacctg	ggcaacacag	caagacccca	104820
gctcaaacaa	agaaaaaaag	ttattgaatt	ttttatttct	atggatcatt	ttttgtagtt	104880
tcttattcct	ttcaccttct	attcccactt	ttgatcccat	ctttttattta	tttagttttta	104940
ttaaatgtat	attttgtctga	taattctgtc	atctacagtt	ttttgtggac	ctgactcagc	105000
atttcttttg	ttcttcggat	tcagactggt	ggtaggctgt	gatttttagtg	attttttggcc	105060
gtgaacatgt	ttcttggact	tttgtctgtg	ggaattctct	gtgtactctg	tataaattaa	105120
gttacttcag	gtgttttgca	ttttcttttg	ccatgcacct	ggggcctggg	tcactaccct	105180
tctggtacca	cttaaaactg	aatttttgtc	ttgggtgctc	gtactgatcc	tgtatgagta	105240
caggtttata	cttactgtag	aaatatgggt	tttgattatg	gggtattgtc	ccagatgggt	105300
ctggagtatt	aatatgctct	ctgttaaaact	taatgtgttg	tccctgtaaa	actccaaaat	105360
tctgaattcc	agaatactac	tggcccaaaa	tgtttaagat	aagggcactg	cctgtatttg	105420
tttctgcctc	ccactatttt	ccttagttta	acacaaaact	acctttttta	aaaacatttt	105480
gagagaattc	agtattggga	agagtttcta	acctgtttct	ggaaatggaa	gtccaaagtc	105540
tgtttctgta	attgtttttt	ttttgagatg	gagtctcact	ctgtcaccca	ggctggagt	105600
caatgacgta	ctctcagctc	actgcaacct	ccacctcccg	gggtcaagcg	attctcttgc	105660
ctcagccccc	tgagttagctg	ggattacagg	tgccaccac	catgcctggc	tgatttttgt	105720
atttttagaa	gagatggggt	ttcgccatgt	tggccagggt	ggtcttgaac	tcctgacttt	105780
gtgatctgcc	cacctcagcc	tcccaaagt	ctaggattat	gtttctgtaa	ttgtaataca	105840
tttattgttt	ttagaaactg	tctttgtctt	agtggtaatt	ttcaataaaa	atagaaatag	105900

91/122

cagtggagtt	attaaaagag	cattagttac	atTTTTccct	ttttcattat	cttcaaatat	105960
tatatatagt	aagtttgacc	tttttaaaat	gtatacttgt	atcagtttta	acacatacat	106020
agattcctgt	aactgtcacc	actataaggg	taaagaacag	ttagttccct	cacctttgaa	106080
gtcaagcccc	acctctatcc	caacacttgg	caaccgctga	tctttctccg	tctcaatagc	106140
tttgccctttt	ctcttttttt	ttcttatttt	tttttttgag	acagcgtctt	gctctgtcgc	106200
ccgagctgga	gtgcagtgag	gcaatctcgg	ctcactgcaa	cctccgcctc	ctgggttcaa	106260
gcagttctct	tgccttagcc	tccctagtag	ctgggattat	aggcacgcac	caccacaccc	106320
ggctgatttt	tttgtatttt	tagtagaaat	ggggtttcac	catgttggcc	aggctgggtc	106380
caaactcttg	acctcaagtg	atccacctgc	ctcggcctcc	caaagtgtcg	ggattacagg	106440
cgtgagccac	tgtgcccaat	caggactttt	tttttttaaa	tttacattca	acttgtcatt	106500
tttttcttgt	atggatttgt	ccttcagagt	cacacctaag	agccctttgc	ctaagcaaag	106560
gtcatgaaga	ttttctcata	tgtttccttt	taaaagtatt	gtggttggcc	agggtgccatg	106620
gcttatgcct	gtaatctcag	cactttgaga	agctgaggtag	ggcagattac	gaggtcagga	106680
gatcgagacc	atcctggcta	atgcggtgaa	accccatctc	tactaaaaat	acaaaaaaa	106740
aaaaaaaaatta	gccggggcgtg	gtggcgggca	cctgtagtcc	cagctacttg	agaggttgag	106800
gcaggagaacc	agtgtgaacc	cgggaggtag	agcttgagct	gagccgagat	cgccgacctg	106860
cactccagcc	tgggcaaac	agtgaactc	catctcaaaa	aaaaaaaaaa	agtattatgg	106920
ttttacactt	tacgtttaga	tatatatctt	ttttgagtta	atgtcgtata	agtatgaggg	106980
ttacgtcaga	ttttttgttt	tttgtttatt	tttacatatg	gatgtctagt	tgttctaata	107040
ccatttgttg	aaaagacaac	ctttactcca	ttgaattgcc	tttgtacttt	tgccatatatt	107100
gtctaggcct	gtttttggac	tcctttttct	gtttcatgat	gtgtgtgtct	attcctttgt	107160
taataccaca	tggctttaat	tactgtatag	taagtcttaa	aattgggtaa	tgctggcctt	107220
ataaaacgaa	ttgggaagtt	tttattttta	ctcttatctc	cattttctag	aagagattgt	107280
gtagaattgg	tgtcatttct	tctttagata	tttggttgaa	ttgggaagtg	atgccatctg	107340
ggcctagggt	tttgtttttt	gtgtgtgaga	cagagtctca	cttctgtcac	ccaggttgga	107400
gtgcagtggt	gagatcttgg	cttactgcaa	cctctgcctc	ccaggttcaa	gttatctctc	107460
tgccctcagcc	tcccaaatag	ctgggattac	aagcgtgtgc	caccatgccc	gactaatttt	107520
tgtattttta	atgcagacag	ggtttcacca	tgtagcccaa	gctgggtctg	aacttgtgac	107580
ctcaagtgcct	tagcccacct	tggcctccca	aagtgcttag	attatagatg	tgagccaccg	107640
tgccctggcag	gggcctaggg	ttttcttttt	cagagtattt	taaaactatga	attcagatta	107700
tttaatatagat	ataggactat	ttaagttatc	tgtttcttct	tgagtgaatt	tttactgtag	107760
tttatggcct	ttgagtaatt	aattgtattg	aattgtcaaa	tttatgagcg	tgtaattatt	107820
tatagcattt	cgggtttgta	gtggtatccc	tcttttattc	ctgggtgttg	caatttgtgc	107880
ttgtttttct	ttgtcagatt	gtatagggat	tttatagctc	tttcaaagaa	ctagcttttg	107940
ttttgatttt	tctgttgttt	tgttttcaat	tttattgatt	ttctgctctt	tattatttct	108000
tttctattat	ttctgcttgc	tttgggttta	ttttactctt	ttttttttct	ccaagttgct	108060
taaagtagaa	acttagattt	ctgggttgag	acctttcttt	tctaagataa	gcatttaata	108120
ctgtaaaatt	ccttctaacc	actgcttag	ttacaccccc	acaaattctg	gtattttgaa	108180
ctgagcacia	atgaaatgtt	ctaatttccc	ttgaatctta	ttcttttacc	aatgaattat	108240
ttagaaatat	gttatattagt	ttgcaagcaa	ttggagactt	ttttcctggt	atttttctac	108300
catttatctt	tcatttcat	atattatggt	cagagaatat	attttgaatg	atttcattta	108360
ttaattttta	aaaataacat	taaaaaattt	tttaaaatgt	gaatatacca	catacagtat	108420
aaagattgta	cattctgttt	ttggacagtt	ttctataaat	gtcaagttga	tttagttggt	108480
taatgatggt	gttcagtttt	tctttattct	tgctgatact	ttgtatgcag	ttatatcact	108540
ttattactca	gaagagtgtt	gaactttcca	actacaattt	ttttttccaa	ttttactttc	108600
agctctatct	cttctgggtg	aatgtcctgt	gaggctctgt	tgtaggtgtg	gtacacattc	108660
aggatgatat	gttctaagat	cagaaatata	tttatcatta	tgtaattccc	tctttatggt	108720
aattttcctt	tagtgcttgc	atggcatatc	tgtagtccaa	tttatataga	cactgcagct	108780
ttcatttgat	attttaaagg	ggcttcttgt	tttttccatt	tttttacttt	tgatctacct	108840
ttataattct	tattttattt	tttatttatt	aggcagcata	tagttgggta	gtgttattta	108900
tttatttatt	gcagtggtgc	aatcctggct	tattgagaca	gagttttgct	cttgggtccc	108960
aagctggagt	cctcagcctc	ccaagtatgt	taccacaacc	tccacctcct	gggttgagct	109020
gattctcctg	tatttttagt	agaaacggat	gggattacag	gcacgcgcac	catgcctggc	109080
tgattttttg	caggtgatcc	acctgctttg	tttcaccatg	ttagccaggc	tcgtcttgaa	109140
ctcctgacct	cccggctgag	tcattgttatt	gtgctgggat	gtgctgggat	tacaggcgtg	109200
agccactgca	tttaattattt	taataataaat	tttaattctt	tctcacaata	cagggttttt	109260
gttggtaaat	gggtatttat	aatgtgtaaa	ttagtataaa	ttatttacat	taaatgtaac	109320
tggtgcactg	atattaatat	ctttggattt	tataattatt	ggtagtaata	taattatatt	109380
actcataata	ttgatttccc	cttttttgct	tttttttttt	tttagtatat	gtttttctgt	109440
ttctccctct				ttttaattct	tatttttttt	109500

92/122

tagtatttgt	tgatcattct	tgggtgtttc	ttggagaggg	ggatttggca	gggtcatagg	109560
acaatagtty	aggggaaggtc	agcagataaaa	catgtgaaca	aggtctcttg	ttttcctaga	109620
cagaggaccc	tgcgcccttc	tgcatgtttt	gtgtcccttg	gtacttgaga	ttagggagtg	109680
gtgatgactc	ttaacgagca	tgctgccttc	aagcatctgt	ttaacaaagc	acatcttgca	109740
ccacccttaa	tccatttaac	cctgagtggg	aatagcacat	gtttcagaga	gcaggggggt	109800
gggggttaagg	ttatagatta	acagcatccc	aaggcagaag	aatttttctt	agtacagaac	109860
aaaatggagt	ctccccatgtc	tactttcttc	tacacagaca	cagtaacaat	ctgatctctc	109920
tttcttttcc	ccacatttcc	cccttttcta	ttcgacaaaa	ctgccatcgt	catcatggcc	109980
cgttctcaat	gagctgttgg	gtacacctcc	cagacggggg	ggcagctggg	cagagggggt	110040
cctcacttcc	cagatggggc	agccggggcag	agggcgcccc	cacctcccag	acggggcagt	110100
ggccggggcg	agggcgcccc	cacctccctc	ccggatgggg	cggctggccg	ggcggggggt	110160
gacccccac	ctccctccc	gacggggcg	ctggcgggc	gggggctgac	cccccacctc	110220
cctcccagat	ggggcgggctg	gcccggggcg	gggtgcccc	cacctccctc	ccggacgggg	110280
cggctgccgg	gctgaggggc	tcctcacttc	gcagaccggg	cggctgccgg	gcggaggggg	110340
tcctcacttc	tcagacgggg	cggccggggc	gagacgctcc	tcacctccca	gatgggggtg	110400
cggtcggggc	gagacactcc	tcagttccca	gacggggctg	cggccggggc	gaggcgctcc	110460
tcccattcca	gcagggggcg	cggggcagag	gtgggtccca	catctcagac	gatgggggtg	110520
cgggcagaga	cactcctcac	ttcctagacg	ggatggcagc	cgggaagagg	tgctcctcac	110580
ttcccagacg	ggggcgggcg	tcagaggggc	tcctcacatc	ccagacgatg	ggcggctagg	110640
cagagacgct	cctcacttcc	cggacggggg	ggcgggcggg	cagaggctgc	aatctcggca	110700
ctttggggag	ccaaggcagg	cggctgggaa	gtggaggttg	tagggagctg	agatcacgcc	110760
actgcactcc	agcctgggca	acattgagca	ttgagtgagc	gagactccgt	ctgcaatcct	110820
ggcacctcgg	gaggccgagg	caggcagatc	actcgcggtc	aggagctgga	gaccagcccc	110880
gccaacacag	cgaacccccg	tctccaccaa	aaaatgcaaa	aaccagtcag	gtgtggcggc	110940
gtgcgctgc	actcccaggc	actctgcagg	ctgaggcagg	agaatcaggc	agggaggttg	111000
cagtgaagcc	agatggcggc	agtacagtcc	agcctcggct	ttcacaactt	tggtggcatt	111060
agagggagac	cggggagagg	gagagggaga	cgagggagag	cccctttttt	gctttctttt	111120
ggattatttg	aatttttctt	ttaaatttatt	tatcttactt	atttatttat	ttttttgagt	111180
gattctcctg	ccaagtactc	caagtactcg	ggatggcagg	catgtgccac	tacaccagc	111240
taattttttt	gtatttttag	tagagacagg	gtttcaccat	attggccagg	ctggtcttga	111300
actcttgacc	tcaagtgate	cacctgcctc	ggcctcccaa	agtgtctgga	ttacaggcgt	111360
gagccaccat	gcccgtgcctt	tttctagaat	ttatatattg	agttcttgat	tgtatctttt	111420
tatgtcctgc	ttctctagga	attacaatat	attacaatat	acatactttt	cacagtgtac	111480
tcacatttaa	tattttgtaa	cttcaagtgg	aatgtagaaa	acttaaccac	cataaaaaata	111540
gaactagggg	tgaggttaaa	aaagagagag	aaaagaaatg	taataaagat	ttaataaacac	111600
cgtttttttt	tttttttctc	tttttttttt	gagacagagt	ctctctttct	gttaccaggc	111660
tgaggtcgag	ttggctgctc	ttggctcact	gcaacctccg	cctcctgggt	tcaagtgttt	111720
ctcctgcctc	agcctactga	gtagctggga	ttacaggtgc	gcccagctaa	gcccagctaa	111780
ttttttgtatt	tttagtagag	acggtttcac	tgtgttgccc	aggatgggtc	cgattttctt	111840
accttgtgat	tcgctctcct	cagcctccca	aagtgtctgg	attacaggcg	tgagccaccg	111900
cgcccgctga	agtcttttaa	tatttttttg	acattgcact	ttttctcttt	tccttcttag	111960
attttagtaa	cccaaagtgt	agttttgtta	ttgtttggca	ggttcctgag	gctttcctta	112020
cttcttttaa	tttttttttc	ctgttggtca	gcttcgaaaa	tttctattca	tctgtcttca	112080
aattcactgg	ttctttcccg	ttatttccat	tctgttattg	agtctttgta	gtgaatttta	112140
aattttgttt	attatgtttt	ttagtcttaa	aattttcttt	ttttgtgtat	gtcttatact	112200
ttgtctctga	aaactctatt	tgtttcagga	tccttagagc	ttcttagagc	atggttttag	112260
tagctactta	aaatttggtt	tatcatccca	gcataatgtg	cctcttgatt	gtcttttctc	112320
ttgtgagata	atgggatttt	ctggttcttt	atatgacaat	taattttgga	ttgtatcttg	112380
gacagtttga	cttacgttac	atgattctga	atcttgttta	aatcctgtgg	aaaatattga	112440
agtttttgct	ttacaagca	gttgacctag	ttaggttcag	tccacaaatt	ctaagcagca	112500
ttctgtcggc	tctgggtcca	tcacagtttc	agttttgtat	cttatctgct	tatgtgcctt	112560
tctgtgtcca	gtctgggacc	tgcccaatgg	tcaggtccca	aagcctttgt	acacttttag	112620
aagcagggcc	atgcacaccc	agctcacgag	tgcccccggg	agtgcacata	caactcgacg	112680
ttttctggg	ctccttcttt	tctgtgatgt	ccttgacacg	tctgccttc	taagaacctc	112740
cctttatccc	tttctgtgtg	tctggctaga	aagtcagggg	tttagattcc	ctatacttca	112800
gcacacttcc	tgtagctatg	tcaacctctg	tggccacgac	ttcttcttct	tgggactgca	112860
gtttctcttg	tcagaaaagta	ggattctttg	agctgctgtc	attgtgtgtg	tggctgtctc	112920
gatgtgcctt	gggagtcgaa	ggagagaaa	gaacaaaaca	aaacaaccca	ggggatttcc	112980
tccactctct	ttgatccgtg	agagccccct	ttctgttctc	tcagaccaga	aatagagggc	113040
ctgtcttgga	acttcttctt	tgtgcatctg	gtgtgcagtt	tcagcttttg	agtcagggcc	113100

93/122

aggagggtgct	ggacaaactt	gtcaggagta	cggagggtact	gcaagttctg	attacttttc	113160
tcagtcacc	tgcttccaag	tccttggatg	catttgtcca	ttgttttgag	ttgcattcca	113220
tgggagagac	agaagagtgt	gcttatttca	tcttgacata	cttatttagga	tttcatatca	113280
aatcaacgga	tgatattctc	tatattaatt	tgctgttttc	ccttttagcaa	gcacattagg	113340
aaaataacac	tttaacaccc	gcctttgggtg	gtttctgtca	taattattaa	tacttgactt	113400
tttttttttt	tttgagacgg	agtctcactc	tgctctttga	ggcattgtcc	ccataaactt	113460
ttggtaaagc	atcaataatt	ttatctttca	tcacacaaag	cttcaccata	aatttgatgt	113520
ttattcttcc	attttagcag	aattcatggt	gctccaatag	gggctgtctt	caaactgatg	113580
ttttctcctt	cttagtgctt	cagagtagat	ctgtgtcaga	tacgttataa	cagggttaata	113640
tgagtttatt	ttgggtgtaa	agtactttga	aattcatgca	tagttttttc	atcatatgca	113700
ttttccatag	ctttgaacac	ccccatgtaa	ctctcctctt	ccacaaacca	aacaatgaaa	113760
aagcaccttt	gtgatggaag	tttattttgc	aataggaact	cacagtgatc	taagccctgc	113820
tattcatgaa	tataattcat	tactggagtc	caagttgctt	tttgggtttt	gaagttctct	113880
tcttcccttg	caggatataga	acaagatgca	gtgaatactt	ttaccaata	tatatctcca	113940
gatgctgcta	aaccaatacc	aattacagaa	gcaatgagaa	atgacatcat	aggtaagcag	114000
tgcttgaaac	tatggcaaaa	aaaaaatgac	aaaaaatgca	cagaactgac	aattttcggt	114060
attgactaag	ataatttttt	cttaacatgg	aattttagcag	ttcccttctt	aatttggttt	114120
ctgagtattt	tttatatcgg	attatagctc	actttaaaag	tttctcggct	gcattcgggtg	114180
cgagggtctt	tgcttgggccc	agatgggctg	cagtgtagcg	gggtgctcagg	cctgcccgtc	114240
gctgagcagc	cgggccggcg	ggcggctacg	ctaaccggca	cagaccaccg	gatggactgg	114300
ccggcagccc	cgcaccagtg	cacgaagtgg	gcgggacaga	aacttctggg	gttgggaagtc	114360
cagtgaggct	aaaagccggt	accaaagtct	ctaggcatca	gggctgcagc	ccaagagctt	114420
cacgaccagt	gggcaactgg	atggccagac	agggtgtctca	gtgggtggcct	ctccgtctca	114480
gggcttcatc	ccacttctca	gtgggcctga	cgctccctggg	caccctggat	gtctacctgc	114540
attagccaga	ggcctgtgac	ttgccttttt	ttgccagttg	attgtgccac	114600	
acacagtgtc	atttctgtgt	catttggcac	agctggaggt	gcaaggagga	gggcagcctc	114660
atgtccagtc	ccagtttcac	gtaactttat	tcttctgaat	aaagacaatt	tgctaacctt	114720
aaaaaaaaaa	aaaaaaaaaa	agttttttctt	atatgtttgga	cccaaattct	taggctttta	114780
cctgaataac	aatgacgca	agatcaataa	atatgacaca	tttattaaac	actcactgtg	114840
tcccagacaa	tattccaagc	actttttatg	gatagactca	ttttaacttc	taaagaactt	114900
tgtgggataa	atacagttat	tttatagatg	aagaaactga	agcacagaga	agttaagtgc	114960
tttgtccagg	gtaacagctc	agatatggga	gagtcaggat	ttgaaactag	accctcacat	115020
accttaactg	cgctgtgtg	tcatactgta	gggtgggacc	aaatatatat	atatactat	115080
tatgccctca	ccccctgcca	aaaaaaaaaa	aaatatatat	ataataatata	115140	
atatatatat	atatatatat	aatatatata	tatataaaat	atatatatat	ataaaatata	115200
tgtattagta	tatatgcata	tatagtatat	attatatatt	agtatatata	ctaatatata	115260
atatacatat	tagtgtgtgt	atataatat	atactagaat	aaaaaaatca	aagtatctca	115320
gagtagtaag	gacaaacatt	tcagaaaaat	gttttcatta	tataacatg	tatgtatgtg	115380
tatgctgatt	caacaaatat	atttcttata	ggttatagca	aaatagtgtg	aaagctttta	115440
ctgtgtttta	tcaggaagac	cttaggtgaa	cgtatattca	cagataaaaag	aggttattta	115500
ttcattcaat	aaatattaca	ttctcataag	tcctaattat	atgtattttt	attcttcaaa	115560
aaagttagta	tttgtgattt	acatgttctt	gggtgctttg	gcacttttag	cagatctgtc	115620
ccgatgttgg	gcttctttta	tccttagtgt	ttcttaaatt	cactcactca	ctgctgggga	115680
cagcaagacc	cctgttagtc	tcagctgtgt	tccattttcc	ggccactgt	accttccagt	115740
tagctattct	ggggtccatg	tcagtgtggc	ggctgttggg	ttttctttct	cccacacaga	115800
tacctataac	taggcctggt	taggcctggt	ggctgttggg	ggcttatccc	tatctgcttg	115860
tatttaaggg	gtactgtttc	actgagtttt	gctgacagat	gttgtcatga	gatttgaggt	115920
tttctgtgtt	gttgctctat	ttttatgtgg	gaatttgcta	ctatcatcat	ccctagacca	115980
gcttttcccta	gtaatacaac	agggatgttc	tgactgatta	gagtttgctt	gtttgaagaa	116040
ttgggtggct	agtgtttttt	ttttgagggg	agtttaatagc	agtttaatagc	ctgactggcg	116100
tgtggataaa	aaggaagcag	tttcaagtca	aataaaacac	ttaaaatgaa	accacactgc	116160
aactctcttt	cttttactta	agcttaatca	aattaatgat	gatgtaatcc	catgaaggaa	116220
aagtcttctg	aaggatcaag	ttgataacat	tttgtgatca	aagaatttga	gaaaacctct	116280
atcccagtgt	ctatcattat	atattttagg	atgttaatta	cctgtgtggc	tttaggcaag	116340
tcatttttcc	tccttgagcc	ccattcttaa	tcctgtccaa	attatttgc	tcctcttgca	116400
gttggactat	tttaatatag	ctgtccttca	agtgagtttt	gttcaaagga	gccttcactt	116460
tagctcttac	tgtgtaccca	ctttgcatag	tcttgtttta	aatgtaatcc	ttggattttt	116520
gggtgtgcta	actaattact	gtttttatgt	gaggatttag	agtgatccag	aatctatact	116580
tgactacct	ccttcatctt	ccacaaatgt	ttgaagtggg	agaattttta	aaaactttga	116640
aggtacagct	gacagaattt	gctgatgggt	tggaagtggg	tggtatgaga	gggaaaaaaa	116700

94/122

ggaataaagc	atgactgcat	tttttgtttg	tttgtttgtt	tgttttttgag	acggagtctc	116760
actctcgcca	ggctggagtg	cagtggcggtg	atcttggtctc	acggcaacct	ccgctccttg	116820
gggtcaagcg	attccccctgc	ctcagcctcc	caagtagctg	ggactacagg	cgctcgccac	116880
cacgcctggc	taattttttt	ttttgtattt	tagtagaaac	ggggtttcac	cggtgtggcc	116940
aggatgggtct	ccatctcctg	acctcatgat	ctactcacct	tggcctccca	aagtgtctgag	117000
gttacaggca	tatatataag	catataaagt	gtgttatagc	atacaaacag	gtatatatat	117060
aaacatgcag	tccacacagc	tgataggaat	gaggcagtag	tgaaggagaa	gttgatgtag	117120
gagaggggac	agttgtttaca	ggaaagaagt	ctggaggcag	aagggatgaa	ttccagtgct	117180
cacatagaag	attgtcttaga	tgggagcaag	gacaattttat	ctagagtcac	aggaaagaat	117240
gcagtacacg	ggtagagatg	caggtgagtt	gaaagatgtg	agagatgatg	gaaataaatt	117300
tctgattgct	tctatattct	caaggaaagca	ggaagcaaaag	tcctcagcaa	agagaataga	117360
agagggtgta	aatatttgag	aaaggagatg	tactgtagaa	aaaaaaaaaa	ctcagttttc	117420
ccttctgaac	tctcacaaaa	cagaaccctt	ccatgactct	agttgtgtgg	gggtttttcc	117480
ctgtcagcta	ccaattctgc	agatgattgt	tcagtgaaca	ccaactgggt	gtcctctaag	117540
tcagttcagt	tctcacactg	tttacctgga	gatagcatca	gatcccacag	attgaggact	117600
ctgtcccaca	agactgcctc	cacttcagat	gccagtctca	agtacaagtt	gtggcctgtg	117660
cttctgactg	accttctata	aattggagtt	cccacagttc	cctccttggg	ttcaataaat	117720
ttgctagagc	agctctcaga	actcagggaa	atgctttaca	tatatattacc	cattttattt	117780
aaaggatatt	acaaaggata	cagattgaac	aggcagatgg	aagagatgca	tgggcaagggt	117840
atgggagagg	ggcacagagc	ttccatgcac	tctccaggtc	atgccaccct	ccaagaacct	117900
ctacagattt	aggtattcag	aagccccct	cccttctctg	tccttttggg	ttttttgtgg	117960
agacttcatt	atataggcat	gattgatcat	tggctattgg	tgatcagctc	aaccttcagc	118020
cccctcatcc	cgggagggtg	gtgggttaggg	ctgaaagtcc	caaacgtgta	attctgcctt	118080
gggtctttctg	gtgattagcc	ctcatcctaa	agctctttag	aggccacagc	cacaagtcac	118140
ctcattagcc	ttcaaaagaa	tccagagatt	ccatgaattt	taggcgctgt	atgctaagaa	118200
actggctaaa	ggccagttgc	aatgtctcag	gcctgtaatc	ccagcacttt	gggaggctga	118260
ggcaggagga	tcgtttcagg	ccatgagatc	aaaaccagcc	tggtcaacat	agtgaagacc	118320
ccttacaaaa	aatttaaaaa	ttggccaggc	gtaatagctc	ttgtctgtag	tctcagctac	118380
tcagaaggct	gaggtacact	gagccctgga	gttgaaggca	gcagtgaacc	atgatcgtgc	118440
cactgactcc	ggcttgggtg	acaaagttag	accttgtctc	agaagaaaaa	ggaaaaaaaa	118500
aaaactgggc	aaagactaaa	taacatattt	cacagtatca	cagatttgta	ttgtctagga	118560
aagtgaatgt	aaacagacca	ggacactagt	atgatccctt	ggtttcatga	aggtcccact	118620
aaagtcattga	acacaaagtg	agactaggca	tcagtgtata	tggtttttcc	agccatgttt	118680
aacagctagc	taaataagcta	attgtttcgc	tgcatgtttat	tttagcagtt	ccttatttta	118740
gcacatttca	tgttttaaaa	tttctaccaa	taacatttta	ataaactttt	ttacagataa	118800
cttcacaaat	ccataatttt	ttaagttaca	atcccagaaa	tagaattgct	cattgaaagg	118860
gtatgttcat	ttttaaagt	atgctagaaa	ctgccaattt	gccttcagaa	aaagggtgtt	118920
gtatccccac	taacactagt	gttagttttc	ttgtgcccct	gctcaagtat	acataattt	118980
aaaaacaatg	ttgggcccagt	ttactagata	aaagggtgtg	tgccctcctta	ttctaattcta	119040
tttgattact	agttagtatg	tatgtctttt	cacgttgggtc	attttatgtt	tggtcccttg	119100
tggattgtca	tgctctttgc	tcatttttct	tttggaaatc	ttcttagtag	tttataagag	119160
ctcttggtat	tttaatgata	gtaacctttt	aactgtcatg	catgctgcaa	atcttttttc	119220
tgtttgtttg	cctttgtatt	ttgttttttg	aggggtttcta	tgtataggaa	ttaaatttta	119280
tgttgttaaa	tcttttgatt	tctgtctttg	catatgtact	tcaaaagact	ttctatttta	119340
agatcaagtg	ttacctgtat	tttcttttag	ttctatttta	aacctcttaa	tttatatgcc	119400
tgtgctgtta	actcccaagt	tgattcacaa	gtgtgtatac	atagtttgaa	tttagtggca	119460
atttaattat	ttacaacttc	ttttgcagca	aggattttgtg	gagaagatgg	acagggtggat	119520
cccaactgtt	tcgtttttggc	acagtccata	gtcttttagtg	caatggagca	agagtaagtt	119580
agttcatatt	ttcacattgt	gcattcctagg	gaatttgggt	tcattgttag	gaatgggctt	119640
cactcagcta	aaaacaaagt	atttttgaga	atttaaatat	tttggatatt	tacaagatca	119700
tataaagcat	actctatctt	gggttaacagt	ttcttttaaa	tataaattat	gtgaactctt	119760
aaaattttca	ttttcatttt	caatgttaat	atttcctaag	ttaaaataat	ttgttttttag	119820
ttctgaaata	atttggggag	tgattgagtc	tgtagtgtat	atgactatta	gaattgggtt	119880
attttatttaa	ataatgcag	tcttcagatg	gctctcctaa	tttgtagatt	aggctttaag	119940
ctaaatggat	gctatataac	taaataccaca	tagattttgt	gaaatggctc	cagaggtttt	120000
ttagattttat	tactgctatg	tgcccttaaa	aaaaatctat	tcattctttc	acttaacatt	120060
tatcagaaga	gtgctctgtg	taagacgtgg	ttaggcatag	tgccagtctt	gaagggaagt	120120
acagcctaata	aaaagacata	gggcatgttg	tttgggtact	gtaatatgaa	gtggcatgtg	120180
ttaaatgtca	ggggagaact	acaaagtcac	aaaaaggtgg	gagagattac	atacaggtaa	120240
aggaatcagg	aatgacacca	tggggagtaa	ggtagtgttg	acctaggcct	ttaaagataca	120300

95/122

atagggacag	tatggaaaga	gtatatTTTT	cccacttaaa	ctcttttctt	ggctcgttccc	120360
tcaaattttc	ccttttgtcc	atgtgcaggc	acttttagtga	gtttctgcga	agtcaccatt	120420
tctgtaaata	ccagattgaa	gtgctgacca	gtggaactgt	ttacctggct	gacatttctt	120480
tctgtgagtc	agccctcttt	tatttctctg	aggtaaagtc	tgcatttctt	ttcacactct	120540
attcgagcat	tccagcctct	aactatcaat	gctggggccc	tgtctatagg	aaataacaca	120600
gaagagccaa	gtcatttcca	aaaagatgta	tcattgtttc	aagttgtttc	tgatggcaag	120660
agtaatttaa	taatatatta	gagagaacat	gaaaattcaa	tgtattaaat	aactctaatt	120720
ttgagaaacc	taattaaact	actgcatgta	agagagtgc	tgtttttaat	tatttggagc	120780
tattttaaaa	ccacagaatt	tgaactttgc	ttccagtgc	ttaaattgcag	accagacttc	120840
agaagagaaa	aaaagtagta	aattttttct	tatgctcatc	atttttactt	tagtcaactg	120900
ataggattgc	ccagtgaaga	agcattttgca	acagacaatg	agtatattaa	tctttttgag	120960
gcatacagtt	tagtataatg	ctctttgtta	ggcttcaaca	agtgaattta	ttttgttggg	121020
aagcaaatga	ctattaagta	gaaagaggat	tcccagtctc	acaaagcagt	aatttagaca	121080
ctcgattctg	cctctttaca	agaatacagg	tactcagttg	atttgttttc	tcactccctt	121140
tctttgctat	aagtttaaat	caacaattttg	tttaggttaa	tatgtcctca	tgggaatggtg	121200
gaaatgatca	gatataaaat	atttggtttg	gttagtttac	tcttttatatg	tttgctggca	121260
aggaaccaca	aatccagttt	agtataattt	ttcactatgt	tcactaaaag	tttgcaccca	121320
gctgtgtagg	tagtgtttgt	ttcttgttaa	cttttttttc	gtctaaaaga	atactttaaa	121380
acttttcaat	ctcaaataac	tgttaacttgc	tgacaggtgt	taacagaaga	agtagatctt	121440
tttgtttttt	gcttatgacc	tgtattttaa	tatttgagct	tatagattag	agattgtgag	121500
agaaatctgt	ttatagtctt	attttccctt	gtgtattttt	tcttcctagt	acatggaaaa	121560
agaggatgca	gtgaatatct	tacaattctg	gttggcagca	gataaacttcc	agtcctagct	121620
tgctgccaac	aagggccaat	atgatggaca	ggaggcacag	aatgatgcca	tgattttata	121680
tgacaagtga	gttatattga	tagatggatt	cagcagatag	ttattgaaca	tttgatatgt	121740
tttgtggaaa	taaagatgaa	taaactcagt	ctctgttgtc	aaggagctca	caggaggcag	121800
cataaaagct	gcttttatat	gggtgtttgt	aaactttggg	ggttcttaga	acaaaagttt	121860
ctgctgggaa	aggggaggtg	tatgtggggt	aaacaggatg	gcaatggtgg	tgttcaagga	121920
gtgtttccca	gaagagagat	tttgtttgga	tcccaaagaa	agaaggggaat	tttgctaccc	121980
agagaagaga	gaaaacaaca	ttctaggcaa	aggcattggc	ccagaagcca	tggaaaacgta	122040
ggggaaagtg	gcactttcaa	gaaacttgag	tttagataat	caaaggagtg	gggaataaat	122100
atgaggatgc	tggtactaat	tggaatagat	tgtaaggggac	cttgaatgcc	tatttatggg	122160
tatattatac	tttctgtata	aatctgctca	ggcacgttgt	taattagttt	tttattagtt	122220
ttcactgaaa	atgagaggat	ggaacatca	tacagtaaac	aaaattgaaa	atatctgggtc	122280
aggcagatga	tgagcttgtg	gccagctctg	taacgttatgg	tattcttttc	atttaacttt	122340
tcttactctg	taaaaaaagt	aattcgtggt	cgggcacggg	ggctcactcc	tgtaatcaca	122400
acactttgag	aggcagaggg	aggtgaatcg	cttgagccca	ggaatttgag	accagcctgg	122460
gcaacatggc	aaaaccggcc	tttactaaaa	atacaaaaat	tagctgagcg	tgatggcggtg	122520
cgcctgttgt	cctagctact	taggggctctg	aggcagaagg	atcacctgag	ccttggggagg	122580
tcgaggctgc	agttagctgt	gatccactgt	actccacctt	gggcagggca	gtagagttag	122640
accctgtctc	caaaaaaaaa	aaaaacaaca	aaggtaattt	gttatttgta	tccttaagca	122700
aatgctaaag	gggtaacttg	gggatagaga	aaagtccaca	gatgttaggg	tttgaagaca	122760
ctaatagtat	ctaggccagt	ggttcctgaa	cattagtctg	tgggctcttg	ctgggctgtc	122820
tgcataggaa	tcacctgaga	gcttattaaa	aataggtttt	caggctgggt	gcgggtggctc	122880
acgcctataa	tcccagcact	ttgggaggct	gaggcaggcg	gattacttga	gggtcaggcgt	122940
tcaagaccag	cctggccaac	atggtaaaac	cccgtctcta	ctaaaaatac	aagaatttagc	123000
caggcatgat	ggcacacacc	tgtaatccca	gctactcagg	aggctgagga	aggagaattg	123060
ctcgagcccg	ggaggtggag	gttgcagtga	gcggagatca	tgccactgca	ctccaggctg	123120
gctgacagag	ggagactctg	tctcagaaaa	aaaaaaaaaa	atagggtttt	agtctgggta	123180
ccggtggctc	acacctgtaa	tcccagcact	ttgggaggcc	aaggcaggca	gatcacttga	123240
ggtcaggagt	ttgagaactg	cctggccaac	atagtgaac	cttgtctcta	ctagaaacta	123300
caaaaaatta	actgggcatt	ttgacgggtg	cctataatcc	cagctactag	ggaggctgag	123360
gcaggagaat	tgcttgaacc	cgggaggcag	aggactgcat	ctcaaaaaaa	aaaaaaaaaa	123420
aaaggtttcc	agtccccctg	tctcagaaat	tctgattctg	caggtttgag	gtgtgaccag	123480
gaatctttat	ttttagaaga	cataccagat	aattctgata	aatagccagt	ttagggatgt	123540
agtctaattt	tcctattttg	caagtaagga	aaataaggcc	cagagaggta	atgattttct	123600
caaagtcaca	gaacaagtta	gtggcagaat	ttggactgga	atgcagttct	taatgttctg	123660
tccagtgttt	attctgggtac	agtatgtttg	tagaagggtat	tacgtaagaa	acattgttat	123720
atagatgttg	agataggaag	agtttacatt	tgaaaatttg	gtctaaaatg	cctgaacatt	123780
caagtcgtgg	aggagtattg	accaacttac	tgaatacaac	ataggagatt	cacattttgt	123840
tacaaaaatg	ctgattttaaa	aggagagttt	tctttttttt	cttctttttt	attttttgag	123900

96/122

atggagtcctt	gctctgtcac	ccaggctaga	gtgcagtgac	acgatctcag	ctcactgcaa	123960
cctccacctc	ctgggttcaa	gcgggttctcc	tgcctcagcc	tcctgagtag	ctgggattac	124020
aggtgggggc	caccacgccc	agctaatttt	tgtattttta	gtagagacag	ggtttcacca	124080
tggtggccag	gccgggtcttg	aactcctgac	ctcaagtgat	ccaccacca	ctgcctccca	124140
aagtgtctggg	attataggcg	tgagccactg	tgcccagcct	gcttggtttt	gtatcatata	124200
tatgcatcat	cataatcatg	cattatcaac	ctttgtattt	ctgtcaggac	atagaaacca	124260
ttagagtgc	tggaagagag	cctttttttt	tttctcgcat	ttaatgcttt	ttttgggtatt	124320
catttcataa	tcagacttacc	aaaacattac	ctgcattata	ccccatcaag	gtagaaatct	124380
ttgtgttatc	aatattgggt	actccctttc	cacaccgagt	catcagtaag	tcctgttcta	124440
tccaaatagg	tcatatgcat	ctagctcacc	cctcagtgct	gttttggttt	gaatttgtac	124500
atgtttactc	ctgatgcctt	gtagttatga	tgatgtgttc	ttattttatt	ctgtgcatac	124560
aagtctctcag	ctcgcttttt	agggaaaaatg	accatgtctt	cctttcctat	aaattccttt	124620
ctatctatca	agtcctcaac	agagaatagg	taccataaaa	tatgtgattg	ttagtttctt	124680
tgctcagtt	gtagtctgat	ccttacagct	tttaaacac	agtagagttc	accgtcaaga	124740
actaaggatg	gttggcaggc	agatagaaag	gtagcaagtt	gacccaacta	tctctgggga	124800
agtgggaaca	aagaaagggt	acatcagcag	tgctctaca	tagctctata	gttctaggcc	124860
tgagggtctca	atcaagtagc	cttgataaag	attctctgga	ggagggtgctg	aaagtgtctt	124920
atacttgcta	tggaatttga	ttttactctg	gatattcttt	taccataggt	acttctccct	124980
ccaagccaca	catcctcttg	gattttgatga	tgttgtacga	ttagaaattg	aatccaatat	125040
ctgcagggaa	ggtggggcac	tcccaactg	tttcacaact	ccattacgtc	aggcctggag	125100
aaccatggag	aaggtaaccc	agaacttcaa	acgtatcaaa	ctacaagaag	ttttattggt	125160
agaactcata	aaatataagg	tggaacaaac	aagcagaata	gcacagtggg	aattgaagca	125220
gtccagcaaa	gtgattaaga	gcagaggcct	tgagtctggc	ctggtatgta	cagtcacgtg	125280
ccacataaca	ttttagtcaa	cagtggactg	cggttagcat	ggtcctgtac	gattataatg	125340
gatcaaaagct	ggtagtgcac	taataacaaa	agttgataaaa	aataaaattt	aataagtaaa	125400
aaagaaaaaa	gaaaaactaa	aaagataaaa	gaataaccaa	gaacaaaaca	aaaaaaatta	125460
taatggagct	gaaaaatctc	tggtgcctca	tattttactgt	actatacttt	taatcattat	125520
tttagagtgc	tccttctact	tactaagaaa	acagttaact	gtaaaacagc	ttcagacagg	125580
tccttcagga	ggtttccaga	aggaggcatt	agttcacaag	gagatgacgg	ctccatgcgt	125640
gttactgccc	ctgaagacct	tccagtggga	caagatgtgg	aggtgaaaga	aagtgttatt	125700
gatgatcctg	accctgtgta	ggcttaggct	aatgtgggtg	tttgtcttag	tttttaacaa	125760
acaaatttaa	aaagaaaaaa	aaaattaaaa	atagaaaaaa	gcttataaaa	taaggatata	125820
atgaaaaatat	ttttgtacag	ctgtatatgt	ttgtgtttta	agctgttatg	acaacagagt	125880
caaaaagcta	aaaaaagtaa	aacagttaaa	aagttacagt	aagctaattt	attattaaag	125940
aaaaaaattt	taataaaatt	tagtgtagcc	taagtgtaca	gtgtaagtct	acagttagtgt	126000
acaataatgt	gctaggcctt	cacattcact	taccactcac	tcgctgactc	acccagagca	126060
acttccagtc	ttgcaagctc	cattcatggt	aagtgcctca	tacagatgta	ccatttttta	126120
tctttttatac	tgtattttta	ctgtgccttt	tctgtatttg	tgtttaaaata	cacaaattct	126180
taccattgca	atagtggcct	acgatattca	ttatagtaac	atgtgataca	ggtttgtagc	126240
ccaaaagcaa	taggttgtac	catatagcca	aggggtgtag	tagggccatac	catctagggt	126300
tgtataagta	cactctgtga	tgtagcaca	atggcaagca	gcctaacgga	aattctgttt	126360
attgattgat	tgattgattg	attgattgag	acagagtttc	actccattgt	ccaggctgga	126420
gtgcagttgc	acagtcttgg	cacactgcaa	cttctgcctc	ccagggtcaa	ccaattatcc	126480
tgccctcatcc	tcccaagtag	ctgggattac	aggcaggcac	caccatacct	ggctaatttt	126540
tgtatttttag	tagagacagg	gtttcaccat	tttggccagg	ctgttctcga	actcctgacc	126600
ttaagtgatc	tgctgtcttt	ggcctccgaa	agtgtctggga	ttacaggcat	gagctaccat	126660
gcctgggcag	taactgaaat	tctctaatac	cattttcctt	atctgtaaaag	tgacgataat	126720
atgcacgttt	acctcaaagt	tactttgatg	attaaagtaa	ggtaattgtat	ataaaatata	126780
tattaacata	gtacctgaca	catggtgaag	atcaaaaaat	gttaactact	tttattacta	126840
ttattattac	gtatttttaa	ataattagag	agcagtatca	aaaattagct	gggcgtagtg	126900
gcatgcacct	atagtccag	ctactcagga	ggctgaagct	ggaggattgc	atgagcctgg	126960
gaattaaagg	ctgcagttag	ccgtgttcat	gccccctgcac	tccagccttg	gtgacagagc	127020
aagaccctgt	cttgaacaa	taaagaaggc	attatgccgc	aacgttagct	tagaaatgat	127080
ccacatatat	caccagtaac	tgtcaacagg	attggaaccc	tagttttggg	tattatgatc	127140
acaaggtatt	attaatagct	tattaataat	aaagcgttgg	ctaggcacgg	cgactcacat	127200
ctgtaatccc	agcactttgg	gaggccgagg	tgggtggatc	acctgaggtc	aggagtttga	127260
gaccagcctg	accaactatg	agaaacccca	tctctactaa	aaatacaaaa	ttagccgggc	127320
gtgggtgggtg	atgccctgaa	tcccagctac	ttaggaggtc	gaggcaggaa	aatctcttga	127380
acccggggagg	cagaggttgc	agtgagctga	gatcgaccca	ttgactcca	gcctgggcaa	127440
caagagcaaaa	actccgtctc	aaaaatataa	ttataataaa	taaataaaaag	taaagtattg	127500

97/122

atgtttgtga	atgattttatt	cttctaata	actagaggag	atttttccag	gaatttcaga	127560
gccagtggg	ttatgttgct	tgtatgtgtc	atgtgtatcc	aggtgaaaaa	acttaattaa	127620
acgctattat	ataataccat	acataaaaa	tgaatttttag	gaatactgaa	gaatgacata	127680
tagaagtc	aatcattaa	agctagtagt	aaacagaata	gagtgtcagc	tgttacccaa	127740
tgatgataat	attttcacga	ttaaaattaa	accttttctg	attttaaagg	aaaagttcag	127800
atctgtatca	tataaagaat	gtaaattttc	agggtaataa	aattaaaatg	cagagagaaa	127860
aatgcaaaaa	tagttcttac	tagatgtgtg	tatgtaagga	acttagacta	attttaagaa	127920
cactgtcaag	accattggtag	ttaggttagga	aaaaagacat	gaatgattca	ttcaacaaaa	127980
actttgagta	tttctgtgct	agatggtagt	gttacagtgg	taaacaaaat	aaatgtgttt	128040
ctgctatcct	ggagcttagt	ctacaaaaaa	gggtacatatt	ggccgggcac	ggtggctcac	128100
gcctgtaatc	ctagcacttt	ggaagatcga	ggcgggtgga	tcacctgagg	tcaggagttc	128160
aagaccagct	tggccaacat	ggcgaaaccc	cgctctact	aaaaatacaa	aaattaactg	128220
gggtgtgggg	cggacacctg	taatcccagc	tactcgggag	gctgaggcag	gagaatcact	128280
tgaacctggg	agacagaggt	tccagttagt	cgagatcatg	ccactgcatt	ccagcccggtg	128340
ggacaaaagc	gaaaatacgt	ctcaaaaaaa	caaaaacaaa	caacaaaggc	acgtattaaa	128400
tacgaacata	aattatttag	aattatactg	aataagttct	catgtttatt	atgtgtgtgt	128460
ccagttacaa	acttttctct	cgtagaatta	gaaatataaa	taataaacat	gagaactcact	128520
tcagtataat	taataattat	taaatgtaaa	taaaaacatc	tatgtacaat	taggcattta	128580
tttaagaatt	atttgaaaaa	aaaacaatgt	ggaaacagat	attttgatat	attgctagtgt	128640
attgaatttg	ataattgtct	tttgaagagt	aaagtgaaca	tatatattaa	agttaaaatt	128700
taactcagca	atcacacgcc	tggtgagtta	tcttaaggaa	atcagtttga	aagtaaaatc	128760
aatatatgca	caaagacttt	aacattttatc	ataaaccaga	aaaatcgagt	ttcaaatatt	128820
atcctatgga	ctattttctg	ctaaaaagta	ttaatatcaa	ctttatgtaa	tactttctgt	128880
acaaaatttt	tggggggagaa	aacccaacaa	aattacatgc	attgtaattt	tttttttttt	128940
ttttttttta	gacagctttg	ctccagcgtg	caggctggag	tgcagtgggt	caatctcgcc	129000
tcactgcaac	ctccatctcc	cagggttcaag	caattctcct	gcctcaggcc	tcccagtagt	129060
ctgggattac	aggcgctcac	caccatgcct	agctaatttt	tatagttttt	agtagagatg	129120
gggtttcatc	atgttgggca	ggctggctct	gaactcctgt	tctcaagtga	tccgtctgcc	129180
tcggcctcct	agactgctga	gattacaggt	gtaagccact	gcacccagcc	ttatgcatta	129240
taattttta	ttgtaaaactg	tacaaaggga	taatacttgt	agtacaacaa	gaagtaaaaa	129300
catttgttat	aggtagttaa	catttgtaac	cagtagaatt	ataggtaaaa	tttattttatt	129360
taaaacagtt	ttagttggat	ttgatttcaa	ctttaaaata	atgcttttca	tctctatcag	129420
gtctttttgc	ctggcttttt	gtccagcaat	ctttattata	aatatttgaa	tgatctcatc	129480
cattcggttc	gaggagatga	atcttctggg	gggaacgtgt	cgctgactgc	tccgtggtct	129540
gttgggccctc	ctgatgagtc	tcacccaggg	agttctgaca	gctctgcgtc	tcaggatttg	129600
actgattgcg	tctgccatta	gggagaaaa	catcacatc	ctttccttca	catcccagta	129660
acagatccca	ttatttgtaa	attttaagtt	gtggaaaaaa	aagataaaa	ccaggcacag	129720
tggcctgtgc	ctgtaatccc	agcacttttg	gaggctgcgg	tgggcgggat	acacggagtc	129780
aggaattcga	gaccagcctg	gccgacatgg	tgaaccccca	tctctactaa	aaatacaaaa	129840
attagccggg	catggtggca	ggcacctgta	atcctagcta	cttggggagg	tgaggcagga	129900
gaatcgcttg	aaccaggag	gcagaggttg	caatgaacca	aaatcacgcc	actgcactcc	129960
agcctgggtg	acaaagttag	actgtgtctc	aaaaaaaaaa	aaaaaagaga	gaaataaaat	130020
tagcctactt	actatcttct	aatcaaaagca	tttgtggtaa	cttaaaatat	actgtattgt	130080
aaagtatcat	gctgtttcat	ttaggccatt	attctatttg	aatctgtggc	tggttctctt	130140
aataaatcaa	gtaatatgga	atatattcat	agcctctgaa	gagctcttta	tgtaagtatt	130200
tatttaggat	actttttgta	aaataagtga	atgaattctt	aggtctcctt	ttttttctt	130260
ttcttgagac	agggctctct	cgctgcaacc	tggaaattct	gggctcaaat	aatccaccca	130320
ccacagcctc	ctgaatagct	gggactagag	gcatgcacca	ccacgcctgg	ctaatttgaa	130380
attttttttt	ggccaggcat	gatggttcac	gcctgtaate	ccagcacttt	gggagaccga	130440
ggcaggcaga	tcacgaggtc	gggagatgga	gaccagcctg	gccaacgtgg	tgaaaccccg	130500
tctctactaa	aaatacaaaa	attagctggg	tatgggtggc	catgcctgta	atcccagcta	130560
cttgggaggc	tgaggcagga	gaatggcttc	aaccaggggag	tcggaggttg	cagtgaagccg	130620
agatcacgcc	actgcactcc	tgcattggtg	cagagttaga	ctccatctca	aaaaaaattt	130680
tttttttaaa	tgatggagtc	tgtctgtgtt	gctcaggctg	gtcttgaacc	cctgacctca	130740
aatgccgcct	gcttcagcct	aagtttcttt	tttttttgta	aagagacagg	gtcttgcctat	130800
gttggccagg	gtagtctcaa	actcctggct	tcaagcagtc	ctccacacct	ggcctctcaa	130860
agtgtggga	ttacaggcgt	gaaccactac	ctataatgtt	gtgtttcact	caaggccttt	130920
tgatttcggt	ttgcattacc	gtgccactac	gtgcatttcc	ttgacctttt	ttgggttttt	130980
tggagtgtct	tcatatgtta	aaccatacct	gattctcttc	aaaatcacac	aaagttagat	131040
atcctaagac	aagaaatcta	aggaggcata	aagaagttaa	ctggttttat	taaactcaca	131100

98/122

cagtaaatga	tagagccaga	aatattcccc	ttctagtgtt	cttcaccatc	agcttaatgt	131160
agcataataa	ttttctaatt	actgttgaca	aataaataac	cctttgaatt	ttcaatactg	131220
ggccttgga	aaattttcct	aatttgaag	agagtattat	cgtattgcca	tttacaagc	131280
tctcctgagt	atctttttct	tctgttaagt	ttacctagga	gataaactgc	tgagtatggt	131340
tgccattttg	gttttttgat	atagggttaga	atgtcctggg	tttttttttt	tttttttttg	131400
gtttttgttg	ttgtcattgt	ttgagacagc	atcttgctct	gtcgcccagg	ctggagtgc	131460
atggcacgat	cgtggctcac	tgcaacctcc	acctcccggg	ttcaagcaat	tctcctgcct	131520
cagcttcctg	agtagctggg	attacaggca	tgtgcaacca	cacctggcta	atttttgtgt	131580
ttttagtaga	gaaggggttt	caccatgttg	gtcaggtggg	tattgaactg	ctgacctcat	131640
gatccacctg	cctcggcctc	ccaaagtgt	gggattgcag	gcatgagcca	ctgcacctgg	131700
ctgaatgtct	gttttttgat	taggcactta	agaaaggcct	aggactaac	cataaaatat	131760
atttttatac	ctttgttgga	tactatata	atagaaaact	gcacttatca	taaccttaga	131820
caccttgaag	aatgtttcaca	agcagaacta	acctagtga	cccagcatcc	agatcaaaaa	131880
cagcattatc	agcccctcta	gaagccctct	tgggccccct	ccattcactg	tccttcttgt	131940
caccagggtg	gctactatcc	tgacttttga	tggcatagat	tagcattacc	tggtcttgtc	132000
attttataaa	taaaaccata	ctgtgtatcc	ttttcttgta	cagctttatt	gtgctaattc	132060
acatttacat	catacaattc	agtgggtttt	atgggttcac	agagttaggt	aaccattacc	132120
acatcgattt	tagaacattt	ttttcactcc	agatagaaac	cccctttact	taaactccaa	132180
atccccact	ccaccagccc	taggcagcca	ctagtctact	ttttatctct	atagagacaa	132240
tagatttgct	tattctggag	atttcataaa	catggaaccg	tatattatgt	ggctttttgt	132300
tgccaactgt	ctttcactta	gcactatgtg	ttcaaaagtg	catcatgtta	tccatgtttg	132360
gcatgtatca	gaattttatt	cctcattatg	gccaaatatc	ccattgcaag	gatttatgac	132420
attttatttg	aattgtaccc	tcctttctgc	catttatcaa	taatgctact	gtgaccattt	132480
gtgtacaagt	ttttgtgtgg	atacagggtt	tccttttgtt	tttaaatttg	aggtggagtc	132540
ttgtctgtgc	gcccaggtgc	gagtgcaagt	gcacaactct	ggctcactgc	aacctctgtc	132600
tcctgggttc	aagcagttct	cctgcctcag	ctcccagagt	atctgggact	ataggcagc	132660
accaccacgc	ccagctaatt	ttttagtaga	gatgggggtt	caccatgttg	gccagtctgg	132720
tctcgaaact	ttgacctcaa	gtgatccacc	catctcggcc	tcccaaagtg	ctgggattac	132780
aggggtgagc	gggtgtggcc	gtaactctta	ttcatttctt	ttgttgata	tacataggag	132840
tagaattgct	gagtcaagag	gtaactctta	aactatttga	aaaactgcca	gattgttttc	132900
cgaaaagggt	gcaccatttt	gcaatcccac	cagcagtgtg	tgagttttac	agcttctcca	132960
catttctattg	gaacttatta	tctgtttggc	tgtttttaaa	aatgatagtc	attccaataa	133020
gttctacttc	agtgtgtgtt	ttgcacttct	ctgatgagta	atgatgttga	gcatcttttc	133080
atttgcttat	tggcctttgt	tctagctttg	gaaaaatggt	tattcaaact	ctttggccat	133140
ttttattttt	atttttattt	atttattttt	ttttgagacc	aagtctcact	ctgtcagcca	133200
ggctggagta	caatgggtgtg	gtctcagctc	actgcaacct	ccgcctcctg	tgttcaagtg	133260
attctcctgc	ctcagcctcc	cgagttagct	ggattacatt	tcaggcacct	gccagcatgc	133320
cgggctgatt	tttgtatttt	tactagtgc	aggggtttcac	catgttagcc	aggctgggtca	133380
caaactcctg	acctcaggtg	atctgcctgc	ctaggcttcc	caaagtgtcg	ggattacagg	133440
cgtgagccat	tgggcccagc	ctagattttc	ttttttcttt	ttttttttga	gaaggagtct	133500
tgctcttggt	gcccaggctg	gagtgcaatg	gcacaactct	ggctcactgc	aacctctgcc	133560
tcctgggttc	aagcgatttt	cctgcctcag	ctcccagct	agctgggatt	acagggtgct	133620
accaccacac	ccagctaact	tttgtatttt	ttttagagac	aggggtttcac	catgttggcc	133680
aggctgggtc	caactcctga	cctcaggtga	tccacctgcc	ttggcctccc	gaagtgtctg	133740
gattaccggc	atgagctacc	aggcccagcc	aattttctca	ttatattgcc	caggctgggtc	133800
tcaaactcct	gggttcaagt	gatcctcctg	ccttggcctc	ccaaagtgtg	gggagtacag	133860
gcgtgagcca	ccttgctcag	cccccttgcc	catttttaaa	ttagattgcc	tttttatatt	133920
gagtttcagg	agtcctttat	atattctaga	taaattgtccc	ttatcaaatt	atattatttc	133980
caggtatttt	cttcattctg	tgagttgtct	ttcctctacc	ttttaaaaaa	gggtgggtttt	134040
tgtttggttg	tttgtttgtt	tttttaagat	aaggtctcat	tctgctgccc	aggctggagt	134100
gcagtggcac	aatcacagct	cactgccacc	tcaacttctc	gggccgaagt	gatcctctta	134160
cttcagcctc	ctgaatagct	agggccatag	atacacacta	tcacacccag	cttttttttt	134220
ctgtttgtag	agacagatct	tactgtgttg	cccaagttgg	tctcaaactc	taggctcaaa	134280
gtgattctcc	cacctctgcc	tcccagatg	ctgggattac	aggtgtgagc	cacacgcaac	134340
ctgtcttttc	actattaata	gtgtcttctc	ccttcagcct	cccagtagc	tgggattaca	134400
ggcaccaccc	accatgcctg	gctaattttt	ttgcattttt	agtagagaca	gtgtttcacc	134460
atgttcaccc	ggctgggtct	gaactcctga	cctcaggtga	ttcacctgcc	atggcctccc	134520
aaagtgtctg	gattacaggc	gtgagccact	gcacccggcc	aaaatattgc	cttcttaaca	134580
gtattgtctt	ctaatttgtg	aacatggatg	tatcttcagt	tatttatgtg	ttctttcatt	134640
tcagcagaat	ttttagtatt	tcagagtaga	agcctttcac	ctccttgggt	catttattcc	134700

99/122

tatgttttaa	gttcttttctg	attccattat	aaatagaatt	gttttcttaa	tttcatttttc	134760
agattgtttg	atgagagagc	atagaaatac	aagtgatttt	tacatgttga	tcttgcaact	134820
tcaactttga	taaactctgat	tgttagctct	aatagttttc	ttgtggattc	tttaggattt	134880
tcaatatata	agatcatgtc	atztatggat	agagatagtt	tttttcttgg	ctagaactta	134940
cagagcaatg	atgagtagaa	gtggcagaag	caaaaatctt	tgtcttggtt	cctatctgac	135000
agggaaagct	ttcagtttca	tcatttaata	tgatgttagg	tgtgggtttt	caataaatgc	135060
cttttttcag	attcaggaat	ttccctatca	ttcctgattt	tttaaggctt	tttttttttt	135120
ttaaatcatg	aaagggtggt	gaatattgtc	atgttctttc	tgtatcagta	taaatgatcc	135180
tatggatttt	gggttttatt	ctgttgatgt	gaaatattaa	ttgattttca	gatgttaaac	135240
caaccttgca	tacctgagat	gaatctcact	tggtcatggt	gtataatctt	ttcaatatgc	135300
tgctggattc	cattttactgg	tattttgttg	aagattttgt	atctgaacgc	ttaagataac	135360
atttacactc	tatcagaaat	gaattgacca	taaatgtgag	agtgtatttg	tgggttcttg	135420
atttctctcc	attccaaaga	tagacataca	tccgtctgta	tgtctgtctt	tatgccagta	135480
ccatactctc	ttgattacta	ttgctttgta	ataagttttg	aaatcagaaa	gtataaatga	135540
gattttggta	tctgagtaac	agtcctcata	gaattagttg	ggaaatattc	cctctttatt	135600
ctggtccctc	tttctttttt	gtttaactgt	gtatcttgga	gattgttctt	tctcaacaca	135660
tgagagccgc	tttccctacc	ctcccacccc	aggtctatag	aggtctataa	gtgtctgttc	135720
aattatttta	tttacttaac	ctattactta	gtcggggaca	tttaagcttg	ttatgtcttt	135780
tattttaaac	aatgctgcag	tgaataatct	tgtatataag	tcattttcca	tcaatataag	135840
tctctctgta	actgaatttt	tagaagtggg	atttctaggt	caacctatgg	ctctgtattt	135900
cacaaaaata	ccaattctgg	tttttcttgt	ggaggtgggg	agtaggaggt	agaatgctgg	135960
aggagaactt	gctgtactca	gctggctagt	cattttagaa	aggtttcctt	agcttctttt	136020
tgtcatatgg	cctcaccaag	aatcaaaaac	attcctattt	accctgtaaa	catgggggctt	136080
tactacccaa	gatacatatt	tctgggatga	tgacagcttt	tcatattgaa	gaaataatgc	136140
tgtgagtaca	gcacatttgt	tggaaacttg	gtcgttaaga	atgtcttata	aattcataca	136200
ttatacattt	tatttttattt	tatttttttag	tttttgatac	agagtcttcc	tctgtcggcc	136260
aggccagcgt	gcagtgggtac	aatcttggct	cactgcgacc	tccatctcct	gggctcaagt	136320
gattctcatg	tctcagcctc	cagagtagct	atgggttacag	gcatgcacca	ccatgcccg	136380
ctaatttttt	tatttttagt	agaaaactgg	tttcaccata	tgaccatgc	tggcctcgaa	136440
ctcttgccct	caagtgatcg	gcctgcctca	gcttcccaaa	gtgctgggat	ccttgtattg	136500
ggtaaaaagat	gaatattgag	ggctgcatgg	tggctcatac	ctgtaatccc	agcactttct	136560
gagactgagg	tgggaggagt	cctggagccc	aggaggggtga	ggctgcagtg	agtttgtgatc	136620
gcgccattgc	acttcaacct	aggaattata	ggcttcagtc	actgtgcccc	gcatgtacat	136680
tttaatatgg	tgcttttctc	ttttagctat	agttatgagg	tacatttcag	agtcgtgtt	136740
gttaagcatc	ttaatagtga	tgagggtgag	tgaaagttac	ttctattttca	aacactgaag	136800
aaaattttgt	acaaatctgt	cacattccaa	gcccaggact	gattgtttca	tatacttcta	136860
attttacaat	ttctattgta	gtccagtgtg	aaaaaaggca	gtattaaaaat	actgaaaaat	136920
tttgatgaag	cgataattgt	ggatgaggca	agctgggatac	cagaatcttt	atatcaacgg	136980
acatatgccg	ggtaagctta	gctcatgcct	agaattttta	caagtgtaaa	taactttgca	137040
tcttttaaat	tttttaatta	aattttacat	ttttttctaa	tctattatta	tatgcccaga	137100
actttcactt	agagtgtgca	gtataatgtg	gtgggttaagt	ataaaggctc	tggagtgtact	137160
tcctgggttt	taatcttggc	tctgccattt	attggcagcc	gctaaccctc	tggatctctca	137220
gtttcttcat	ctgtaaaatg	agaataataa	agtgaagaaga	tgccaacatc	atttactctg	137280
ggctgcataa	ctgatacttg	gaaaaagtat	tccttttgagt	ttaagaatta	agttgggttat	137340
tcatttttagc	ttgtaataaaa	aagatagtga	ttcataggat	atgccactta	ctgaaattta	137400
ccacagatcc	aatcataaaa	tcactttctc	ttccctaag	atagcttgat	taacatgtaa	137460
aggtgtgtaa	aggcttgatt	acactaccct	gatccgtacc	ccagttccca	gcagcaccat	137520
gaaaaaggga	tttcaacata	tttaattact	ttcagtagaa	agtaacagtg	gtaggccagg	137580
cgagtggtct	cacacctgta	atcccagcac	tttgggaggg	cgaggtgggc	ggatcacgag	137640
gtcaggagat	tgagaccatc	ctggctaaca	cgatgaaacc	ccgtctctac	taaaaatata	137700
aaaaattagc	cgggcatggt	ggcaggcacc	tgtagtccca	gctacttggg	aggctgagac	137760
aggagaatgg	cgtgagcccg	ggaggcggag	cttgcatgta	gcttagattg	tgccactgca	137820
ctccagcctg	cgagtgagg	cgagactctt	gtctcaaaaa	aaaagaaagt	aacagtggta	137880
ttgggagact	gaggagccta	gaaagtactt	gaaggaagta	aaagggttgt	ttgaccacat	137940
tgtatttgga	aagccagctt	tttcagctgt	gtcagctttg	tgtagtgtat	tttagtctct	138000
cttttagaaa	ataacggaca	aggccgggca	cggtggctca	cgctgtaat	cccaccactt	138060
tgggaggccg	agacggggcg	attacctgat	ctcaggagtt	cgagaccagc	ctgggcaaca	138120
tggtgaaacc	ccgtctctac	taaaatacaa	aaagttagcc	gggcgtgggtg	gcgtgtgcct	138180
gtagtccag	ctactccgga	ggctgaggca	ggagaattgc	ttgaaccggg	gaggcggagg	138240
ttgcagttag	ccaagatcac	accattgcac	tgcagcctgc	gcgacagagt	aagactctgt	138300

100/122

ctcaaaaaat	aataataaaa	taaaaaagaa	tggacagtaa	acctaaatga	gttcattccc	138360
aaagatgatg	ttattcttaa	gggatgggtc	atztatttta	gaccttacat	aaagtctatc	138420
aattgcgtag	tttttcactt	ctgtaattgt	gtgtatgtat	aatgtaaata	tatatgtttt	138480
tgttttggtt	tgggtttttg	agacggagtc	tcgctctggt	gctcaggctg	gaatgcagtg	138540
gtgcaatctc	agctctctgc	aacctctgtc	tcccagggtc	aagcggtttc	tctgcctcat	138600
cctcccaagt	agctgggact	acaggcacgt	gccaccacgc	ccggctaatt	ttttgtattt	138660
ttagtagaga	tgggggtttc	ccgtggttag	caggatggtc	tcaatctcct	gacctcgtga	138720
tccaccggcc	ttggcttccc	aaagtgttgc	tattacaggc	atgagccacc	acaccagca	138780
tgtatttttt	aaatgtataa	aatgaagcag	aaaagagaaa	tgataatttt	tcttcattct	138840
gaaagattat	cttcaccagg	cgcagtggct	cacacttgta	atcccagcac	tttgggaggc	138900
ctcggcaggc	ggctcacttg	agttcgaaac	cagcctggcc	gacatgggtga	aactccgtct	138960
ctactaaaaa	taaaataaata	aagatggttt	taatatatgt	tttagtttta	tgatttttag	139020
atctttctga	aatttttctc	aaggcaagta	aatttgtatc	agttgggtata	ttggtaccca	139080
tctatgaaat	aacttattag	gaagatatct	ctaaaataag	atcactttgc	ctaaaataaa	139140
ctgatataat	gatgttcaca	gaatttttct	tttaaccgac	ttgataaatg	cattattctt	139200
gacgtccagg	gatccacctt	cctcagcctt	ccaaagtgtc	gggattacac	acatgagcca	139260
ccgcacctgg	cattattctt	ataaaagggt	aaatttctag	ttaaagttaa	tgctcctctt	139320
gttcattgtac	cattgtctat	tttcttcctt	tcctactcac	agtaatcatt	cttatgggtat	139380
gcacttttgt	ttgcttattt	ttatgtaatt	gatattacgc	tccattctgt	acgttggtact	139440
ttcattccaca	gtgagttttg	gacattccta	tgttcatcta	tacagactta	cttcatttta	139500
actacactgt	atgtaataat	tactataact	catcactgta	catcactgta	gcagagcatc	139560
tcatagtgtg	tgtattactg	ttttgccatt	ttgggtatcaa	tgagtattta	agtcattttgc	139620
agtttttccc	tcttataccc	agtattacag	aggatctctt	tttatatgct	tctttgtacc	139680
aagaggcaga	ttaaaaaatt	tttttttgaa	aaaatttttg	aaaaaaaatg	aaatgaagtc	139740
tactatgttt	gcccaggctg	gtctcaaatc	ctcaggctca	agcaatcctt	ccatcttggc	139800
ctcccaaagt	gctgggggta	caggcatgag	ccaccatgcc	tggcctacat	tttaaatttt	139860
gatagctctt	acaatttact	ttgtaaagta	tctgcatcat	tttatgttct	caccagtctt	139920
taataagaat	acttcatact	tttggctgga	cacagtggct	cacgcctgta	atcccagcac	139980
tttggggagg	cgaggcgggc	agatcaagag	atcgagacca	ccctggccaa	tatggtgaaa	140040
ccctgtctct	actaaaaata	caaaaattag	ctgggcgtgg	tggcgacccc	gtagtccag	140100
ctactcgaga	ggctgagaca	ggagaatcac	ttgaaccggy	gaggtggagg	ttgcagtga	140160
cttagatcac	accatgacac	tccagcctag	caacagagtg	agactctgtc	tcaaaaaaaa	140220
aaaagaatac	ttcagactta	attttttttc	cagtcctaa	tgtttgctaa	tgagattgag	140280
tttcttttgg	tatgtctctt	gattgttcag	gttttttctt	ttatgaattg	actgttcac	140340
tctttttcac	attatttctg	ttgggtgatt	ttattagtga	cttggttaaaa	ttctgtatat	140400
tttttcagca	tgacacttca	ttattcaaaa	aaaaaaaaag	attctctatg	tttctcgata	140460
ctaatcattg	gttggttaata	ccttaaaaaa	aagaccttta	ctgtattttt	tgcttttttt	140520
tttttttttt	tttttttttt	tttgagatag	agtcttgctc	tgttgcccag	gctggagtg	140580
aatgggtatg	tctcggctct	cagctcactg	caactgcaac	ctctacctcc	ctgtttcaag	140640
caattctcct	gccttagcct	cccaagtagc	tgggattaca	ggcatccacc	accacaccca	140700
gctaattttt	gtatttttag	tagagacagg	gtttcaccat	gttggccagg	ctggtctcaa	140760
actactggcc	tcaagtgtat	cgcctgcctc	ggcatcccaa	agtactggga	ttacaggcat	140820
gagccacagt	gcctagccac	tttttgcttt	ttacttttgt	tttatagtac	tatagtttta	140880
gtataaacag	atgtatgtat	acacacaact	atggctttat	aatatgtttc	agtcattgtt	140940
agagcaaggc	ctaccttttg	ggtgcttctt	ttacaaaatt	gtcttggtta	ttcttgtgcc	141000
ttttttctta	tttgtgaatt	ttagaattgt	gaattacctg	ttgactcacc	atgttttgta	141060
aactgaggat	tttgaatgga	attgcactca	attaaagatt	atcttgcttt	ctgtgcagca	141120
atgttttatt	tcaaataatc	cctactttta	attacttagg	atagctataa	attgtgtttc	141180
tggctttcta	gatttagatg	aaacgcttta	aattgattgt	tttctcctaa	atttaaaact	141240
gattgttaga	agttaaagtc	ttctgttcat	tcttatttag	gaagatgaca	tttgggaagag	141300
tcagtgactt	ggggcaattc	atccgagaat	ctgagcctga	acctgatgta	aggaaatcaa	141360
aaggtttgtg	gtgtttttat	acttcatatt	aagcctttac	tcacattagt	gattgactgt	141420
aagtcaaaga	ccacttaagg	tttaaactgt	ttattttgta	aagtaaccac	tgtatctttc	141480
acctgtgtgt	tatagtcaga	agggcttctt	cttcttccct	gtagtccat	ctttatgcaa	141540
tctcctctga	atcaaaagtt	agtgaacttg	ctttgccact	ccagaaggca	catgaattatg	141600
aaaaagcatt	gtctattttc	ttatttaatt	gcaaaatacc	cgacctaaagt	tggacttaatt	141660
gtttgagacc	gtttatttta	tttaattata	ttttttctct	tttctttttt	ttttttgaga	141720
cagttcttgc	tctgtcaccc	agaccggagt	gcagtggtct	gaccgcacct	cactgcaacc	141780
tctgtctcct	aggttcaagc	gatttttctg	ccctatctct	ctgagttagct	gggactacaa	141840
gtgcgcacca	ccacacctgg	ctaatttttg	tatttttagc	agagatgagg	tttcaccacg	141900

101/122

ttggctaggg	tggtctcata	ctcctgacct	caagcaatcc	atccgccttg	gcttcccaaa	141960
gtgctgggat	tacaagtgtg	agccaccatg	cctggcctta	ttaaattatt	tttattaaat	142020
ttcctcaaga	ttgatgaaag	taatgaaata	taaaagtaat	gaaatatatg	tggaaaatag	142080
actggattaa	gaaaatgtgg	cacatatata	ccatggatac	tatgcagcca	taaaaaagga	142140
tgagttcatg	tcctttgtag	ggacatggat	gaagctggaa	accatcattc	tgagcaaaact	142200
gtctcaagga	tagaaaacca	aacaccgcat	gctctcactc	ataggtggga	attgaacaat	142260
gagaacactt	ggacacaggg	tggggaacat	cacacgctgg	ggcctgtcgt	gggggtggggg	142320
gctgggggag	gaatagcatt	aggagatata	cctaataata	atgacgagtt	aatgggtgca	142380
gcacaccaac	atggtacatg	tatacatatg	taacaaaagt	gcacgtttgt	cacatgtacc	142440
ctagaactta	aagtataata	aatttaaaaa	aaataaatat	atgtggaaaa	tattaatagg	142500
tcaaaattca	aattgttcat	ttaatcagaa	gagtagttta	gtcaaatcca	agggttagac	142560
aacagaaatc	ttttttgtca	agtgcattct	ttgtgactga	tttcattttc	ttcctggttt	142620
acacaggaag	atcttcagaaa	caaagtgtga	tccgtgacag	atggtatcta	gaagttttta	142680
gtttgggtga	attgacagta	ttttattgag	taaaagatac	taatttttgt	aagaagaaaa	142740
attcaatttt	gataagtatg	tttaagatta	agagctattg	gccaggcgct	gtggctcatg	142800
cctgtaatcc	tagcactttg	ggaagctgga	gcaggtgggt	cacgaggtca	agagattgag	142860
accatcctgg	ccaacatggt	gaaaccctgt	ctctactaaa	ttagccaggc	gtggtggcac	142920
atgcctgtgc	acccgcctcc	gggtttaagc	gacccactcg	cctcaggctc	ctgagtagct	142980
gggattacag	gcgcctatgg	taatttttgc	attttttagta	gagacagggg	ttcactacat	143040
tggccaggct	ggtctgggtc	caaactcctg	acctcagggtg	atctgcccgc	cttagcctcc	143100
caaagtgtcg	ggattacagg	catgattcac	catgtctggc	catttatctt	attttctttt	143160
tttttttttt	ttttgtttga	gacggagtct	tgctgtgtcg	cccagagctg	gagtgcaatg	143220
gtgcgatctc	agctcactgc	aacctctgcc	tcctgggttc	aagcaattct	cctgcctcag	143280
tcttccaagt	agctgggatt	acaggcgctg	gccaccacat	ctagctaatt	tttgtaattt	143340
tagtagagac	agggtttcac	catgttgggc	aggctgggtc	cggaaactcct	gacctcgtaa	143400
tctgccacc	tcggcctccc	aaagtgtgta	gattacaagt	gtgagccact	gtgccagcc	143460
atcttatttt	ctttcttttt	ttttgtcggg	tgggaggggg	acagagtcta	gctctgtcgc	143520
caggcttggc	tcactgcaac	ctctgcccc	cagggttctag	caattattct	gcctcagcct	143580
cccaagtagc	tgggattata	ggcacctgcc	accacgcctg	gctaattttt	tgtatttttt	143640
agtagagatg	gggttttgct	atgttgacca	tgtgtgcctc	aagtgatccg	cccactttgg	143700
cctcccaaag	tactgggctt	acaggcggtg	gcttgtattg	ggtaaaagaa	caatattggg	143760
ggctgcatgg	tgggttcatac	ctgtaatctg	agcactttgt	gagactgaga	tgggaaggagt	143820
gttggagccc	aggaggggtg	ggctgcggct	gcagtgaatt	gtgatcacgc	cattgcactt	143880
ccacctaggt	aatggagcaa	gacctgtctc	gattacaaca	aaacacaatt	tttttaagga	143940
atactgggaa	gaggtcagtg	gtgggttttag	aacagaggaa	gtgccagatg	acctttgtga	144000
ggcattggcc	aggaagaact	ctacagtgtc	tttaggtagc	ttctgtccat	aaggataatg	144060
gggtctcctc	cccagtatta	atagaaaaatc	tctgagctgt	ttttttttgt	ttgtttgttt	144120
tgtttttttt	tcctgagatg	gagtcctctc	ctgtgcgcc	ggctggagtg	ctgtggcgcg	144180
atcttggctc	actgcaagct	ctgcctccca	gggtcacacc	attctcctgc	ctcagcctcc	144240
caagtagctg	ggactacagg	tgtccaccac	cacgcccagc	taattttttg	ttatttttag	144300
tagagatggg	gtttcaccat	gtcagccagg	atggtctcga	tctcctgacc	tcgtgatccg	144360
ctgcctctg	ccttgcaaag	tgctggagtt	acaggcgtga	gccaccgtgc	ctggcctggt	144420
ttttttgttg	ttgttattta	tttattttatt	tatttttttt	ttgagacaga	ctctcgctct	144480
gtcggccggg	ctggagtgtg	gtggcacgat	gtcggctcac	tgcaagctct	gcctgccagg	144540
ttcaagccat	tctcctgcct	cagcctcctg	agtagcaggg	accacaggcg	ctcgccacca	144600
cgcccggtca	attttttgta	tttttagaag	agacgggggt	tcaccgcatt	agccaggatg	144660
gtctcgatct	cctgatgtcg	tgatccgccc	acctcgccct	cccaaagtgc	tgggattaca	144720
gggtgtgagcc	accgtgcctg	gcctgatttt	tttttttttt	taatctgggtc	tcatacctct	144780
gacagctcat	gaagaagtgc	tcctgcttca	tatgtatatg	tgtagcata	gtgttaacat	144840
agcataggtg	ttcgggtgtt	gcagtttctg	tttgttttat	atgaattaa	gtgtattatg	144900
agcagttgaa	gatatatagg	aaattttttc	ccaaaccact	atctctgctc	gttctattca	144960
ttcagttctg	ttatgttatt	ccttcattca	ttcattttat	agaacagtg	agtgcctact	145020
gtatgcatct	attgttctgg	gtcctgggga	agaaaacaaa	gttcctgctt	tcattggaact	145080
tacattatat	tggcggagac	agtaacagac	aaacaaatgt	agcctgtgta	catgtgttac	145140
atgaaaagca	gggtaggggg	ctgggagaga	gtgctatttt	gtgctatttt	cgaggtgggt	145200
gtcaggaaag	gcctcactga	ggaggtggca	ttttgagtag	acctgagcgc	agcggggggc	145260
taagcccagg	cagcatgtgg	aggaagagtg	ttcttggtga	aaggaaacaag	gatagaggcc	145320
cgaagctaga	gagctcagca	tgatcaagga	acagcaagcc	ccgtgtgggt	ggaatggagt	145380
gagcaaaagga	atgagcagta	gaaggtgagt	gagttgggag	gtcaccagag	accatggcaa	145440
ggacttgaaa	gtgtcagggg	cacattggaa	gttggagcag	ggaaatgatg	ggatttatgt	145500

102/122

ttttatgttt	agtgttttta	agggattgct	ctatcagcta	tttggaaaat	145560
gcttcaagaa	gagaagcaga	gaaacaacat	tcttgccata	gtcatagtct	145620
tgatgggtgt	gtggattagg	ctggtagtgg	aagaccagtc	cagttcgggt	145680
ggtagaggca	aaaagattat	atttctacca	gcaagcccat	ctatgaagtt	145740
ttaatttaat	tgagacatgc	ccacataaac	taataaaatag	gaatttctgc	145800
aacaccctcg	tatatcctgg	ttcttctttt	agttgtccag	atgtctcttt	145860
ttttttgggt	gtgtaggagc	ctagagattg	aattttattca	cccaaaaggc	145920
ttactatgtg	ccaggcacta	tgctgaatgc	caaggatgta	aataagaggg	145980
gtctgtttta	ctccagcttg	gttccctttt	aatgaccctg	acttggttaag	146040
atcctacaga	atgtttaatc	ttctgtactt	tcctgggtgt	gttatttagc	146100
ttccttgaca	tttcttgtaa	actggaagtt	acacctatag	tcttgatgat	146160
catttttagat	tagaacacat	catgtgttgt	atatggtgtt	tttgaaagcc	146220
ttgggtctgta	cattaaaatg	ttgcctgaat	ggatacacat	aaaatttaac	146280
ttagagatga	gaagaaagag	gtgcctttta	cttttcaata	taccttttcc	146340
gaactttctt	gccctatgca	tacgttattg	cttaatcatc	cacctcatct	146400
ggctttctgt	tgcatttgga	atgaaatcta	gcctctttgc	tgttacctgt	146460
tgtctggctc	tatcacctta	ctttgaacca	ctcctttcat	ggactgagct	146520
tatcttttat	tcttttgctg	aagtttcttc	actttgagtg	cctctgcagt	146580
tggctgtggc	aagccctgcc	atggctttca	tgcaaggatg	gttccctcctt	146640
tattatctct	tcagagaggg	accttcccaa	ctccgatgat	ctaaaaatcct	146700
cactcactac	cacttctttc	ttttcttttc	cttttatctt	tttttttttt	146760
gayalagggt	cttgctctgt	tgcccagget	ggaatcacga	ctcactgcag	146820
ttgggtctaa	atgatcctct	cacctcagcc	tctcgagtag	ctggaactgc	146880
caccatactt	ggcttattat	tttacttttt	gtagagacag	ggtttcacca	146940
caagctcctg	ccgcaagcaa	tccacatctc	tcagcctccc	aaagtattgg	147000
gtgagccact	actcctggcc	tattttctta	ttcactgtct	aaaattatct	147060
tttacatact	tgtttatagc	ttatttctca	gctggacatg	gtgcctcaca	147120
caatactttg	ggaggctggg	ttggagaatt	ggttgagccc	aggacttcaa	147180
ggcaacaaa	tgagaccctg	tctataaaaa	attgtttaaa	aattagctgg	147240
acatgcctgt	gggtccagct	acttgggagg	cagagggtggg	agaatcgctt	147300
gggtgaggcg	acgggtgagcc	atgatttgtc	cactgcactc	tagcctagt	147360
accatgtgtc	taaaaagtaa	ataaaaaatg	tttctctttc	atgactagaa	147420
atgtgggcag	ggagtttgtc	tatactattt	ggcactatat	ttcctgattc	147480
cctagcacat	ggtaagtact	ccttaaatat	ttattgactg	aattatttaa	147540
tttcatattg	gattatctga	gtggtaagat	tacggattat	atztatgtaa	147600
ttttttaaac	ttggttgccc	tttgccacac	tgacatagac	actaagtttt	147660
ctacttccga	ggatactcac	agaggccatt	ctcttctcaa	tcccaaaata	147720
tttagcactt	tcaagctaata	gcaattctta	gatgatgtat	ctgtgtatat	147780
ctctacaaat	gtagaaattg	aagtctgggc	acagtggctc	tcacctgtaa	147840
ctgggaggcc	aaggcgagcg	gatcactgag	gacaagagtt	aagaccagcc	147900
ggtaaaagcct	tgccctctatt	aaaaatacaa	caattagggc	cgggctgtgt	147960
ataatccca	gcacgttggg	aggccaaggc	aggcagatca	cgaggctcagg	148020
atccttggt	aacacagtga	aaccccatct	ctactaaaaa	tacaaaaaat	148080
gggtggcacg	cgcttgtagt	cccagctatc	gggaggctga	ggcagggtgaa	148140
cgggaggcg	gaggttgcaa	tgagctgaga	ttgcaccgct	gaactccagc	148200
jagggagact	ctgtctcaaa	aaaaaaaaaa	aaaaacaatt	agccaggcgt	148260
icgagtacct	gtaatcccag	ctactagggg	ggctgaggga	ggagaatcac	148320
jaggtggagt	ttgcagcggg	ctgataatgc	accactacat	tccagcctgg	148380
jagactctgt	cttaaaaaaa	aaaaaaagaa	agaaagaaat	tgaggaatgt	148440
jtctgtgatt	tgtaggaat	cacacagcag	gttagtagca	actacagggc	148500
iataccacct	tgacaatggt	ttgtttacag	ttcggtcccc	cttccctctgc	148560
tccttattg	agggcagctg	gaaagaattt	tcatcattta	ctagcctata	148620
jagttttgaa	accttgataa	tagagcacag	aggaaaagac	tgagttttct	148680
agtcttgct	ctatggccca	ggctggagtg	cagtgcacac	atctcagctg	148740
tgccctcca	ggttcaagca	attctgcctc	agcctctcga	gtagctgaga	148800
jtgtcaccac	gcccagctaa	tttctgttt	ttgtttcggt	ttgttttttt	148860
jtcttgctct	gtcaccaggg	ctggagtgca	gtgggtgcgt	gttggtcac	148920
jtctcctggg	ttcaagcaat	tcttctgcct	cagcctcccc	agtagctggg	148980
gtgccacca	tccctagtgc	atttttgtat	gttttagtaga	gatgggggtt	149040
iccaggctgg	tctcgaactc	ctgatctcag	gtgatctact	cgtctcagtt	149100

103/122

tcccaaaagt	ctgggattat	tggcacacgc	ctatTTTTgt	atttttagta	gagacggggg	149160
ttcaccatgt	tggtagact	ggtctcaaac	ttctgacctc	aagtgatttg	cccgcaccag	149220
cctcccaaaag	tgctgggatt	acaggcgtga	gccaccgtgc	ccagccaaga	ttgagttttg	149280
aaaagagcct	tctgagatta	tgagaagggc	aagcaagata	acttaagaag	ttacattaaa	149340
atcatctaag	agacagtgtg	acaagaagga	attgtaaaat	gatgttatga	gcacgtgtcc	149400
aatgtagtgg	caatcccttg	tgcttcgata	cattggtggg	agacaaaact	gtacttaaat	149460
tgataaatcc	cttacatgtc	attttaagga	gcttagactg	actcccatca	tgtagacatc	149520
agagatttct	tttttttttt	tttttttttt	tttttttttt	tttgtgacag	agttttgtct	149580
ttgttgccga	ggctggagtg	caatggcgtg	atctcggtct	accacaacct	ccacctccca	149640
ggttcaagca	attctcctgc	ctcagcctcc	cgagttagctg	ggattacagc	catgcaccac	149700
cacgcctggc	taatttttga	tttttagtag	agacgggggt	tctccatggt	gtggctgggtc	149760
tcgaactcct	gacctcagtg	gatcctcccg	cctcagccac	ccaaagttct	gaaattacag	149820
gcgtgagcca	ccgcgcccag	cccagagatt	tctaaacaga	gttctaacca	gatgcttttc	149880
cctgtcagta	gaatgagaat	gaattggagg	tgggagagac	tggcatgagg	gacaccagtc	149940
agccagtggg	attagctggg	aatggtgata	ggagaagaaa	aagattcaaa	gttaggtagt	150000
ggtagcaaga	attagaggga	aggtcggatt	tatgatattg	ccaagggtga	attctaaggt	150060
gaaatttggt	ggcagatttc	atgtgtaaat	tgggaaggta	gattgagttt	ttttaacatg	150120
ggttttctaa	catgtcaata	gagtgaactc	gcaggggggc	ctgacgagag	aacagtgcac	150180
ggggtgattc	aacagccagt	tgagccttca	tcgagagcat	ttaacactgt	gactctgtag	150240
actctggttg	gcagtaaaat	ttcattaaac	caatatattaa	acccttaggt	aataataaaa	150300
attgaggggaa	aaggatccag	gttttgattt	ttttatgaat	tcagttattg	aattaacacg	150360
gaccttgccct	caagaaataa	tctaccaaca	attaacttgt	tttaaagcaa	agttaggaag	150420
tgagcatggt	caaattatta	aataaaaaag	taagctgtgt	atttcattca	tagaaataga	150480
ggctggccta	cttcggatga	ttctcagcat	gtgattacag	atgtgggctt	atacatccta	150540
gggagtttaag	gcgtactctg	gcttggatag	agttagagctc	tttgaactc	ttctctcacc	150600
cagctagttt	atatagacta	gagaactaga	atgtagagc	atactctgtc	ttagaagccc	150660
ttttatatag	gagctgggtc	ggaaggtttg	aaaacataac	aaatgtgttg	gtgtctccca	150720
atgtattgct	agattcttac	ccaagagcat	tatcctgggt	agggtttggt	ttgggtttgt	150780
tttggttttaa	aagcagcatg	gcaaatattat	actagattgt	agttctttca	tcaagtgagg	150840
agagtagaag	aaaagtccag	aactctgaaa	caccttttca	aaagtttttc	aagccatgat	150900
gtttgcaagt	taaatgctct	gttatgtaag	caatataatc	agttttttatt	aatgtaacat	150960
tccctagtag	tttgggggat	cacacaaaaa	agaatatcca	tatctggaag	caacagcttt	151020
taataatgag	catttggttg	gtggtgggtg	tagtggtttt	tttttttttt	tttgagttgg	151080
agtctcgctc	tggtgcccag	gttggagtgc	agtggcacga	tctcagctcg	cttcaacctc	151140
tgctcccagg	ttcaagcaat	tcttctgcct	cagcctcctg	agtagctggg	attataggca	151200
cctgctacca	tgccctggctg	atttttatta	tttttagtaga	gacaggtttc	accatgttgg	151260
ccaggctgggt	ctgaactct	taacctcagg	tgaatcaccc	acctcggcct	cccaaagtgc	151320
tgggaattaca	ggcatgaacc	accatggcca	gccaataaag	agcattttta	atgtaaaatt	151380
atgcatgaaa	tgtacattca	attttgtctt	tgtttactag	gatccatggt	ctcacaagct	151440
atgaagaaat	gggtgcaagg	aaatactgat	gaggtaaatc	ctacctttag	gataaaaaaga	151500
tttctgttta	taagtgccac	cctcatgtaa	gtgaggttta	aaattttctc	tttcttttagg	151560
tcccatgttt	cttaacatct	aacaattgcc	tggtagtagc	ccagaatgta	ccaagaaagg	151620
gtggtccctt	ccactgaagg	atatacatgc	attcaagttt	agtgaaggta	tcttcagtca	151680
gaggctagga	tgtagatcaa	aagctcaaat	gtttccttcc	ccatcagcca	gcaggcatca	151740
gtaatcagtg	gcttgctttt	ttaatagtta	attaaagtaca	ccactggcag	ttttacttca	151800
aaaggaaaat	gttttatttt	gaccatctaa	tatgaaagta	ttgagagatg	ggaggtgaaa	151860
tagttgacac	tgggaagacag	ggaatgacat	gttaatatcc	gttcggtgtt	aggtatccag	151920
gttttttctg	acatgggaat	gaaattctta	tccaaataag	atagccagag	ggtggcccag	151980
catttggttaa	gagcactgag	tatgtgcac	tgcatccatc	tagaaattat	gtgcgtaagc	152040
cagtttaaat	tattttcttt	aggcccagga	agagctagct	taatgaataa	ccattatcac	152100
agtcagtgc	attatgcagc	aggctcagta	tgatcaaccg	tggaaagattg	ctaaaatgat	152160
aaggatgact	tcgttttgtg	taaactaaaa	agtattattt	ttagagaaat	ctacaaaggt	152220
agaacataag	gggtttcttt	gcctttgaag	gattaactgc	tccaggtgta	aaaataaaaa	152280
cataagcaac	tagaaaattg	acaaactaaa	tgaaacaact	tgtggggatt	accttcttat	152340
gggcaataca	gggaaaccat	ggaaacccaa	cagagcccag	gtttgcatat	attggacaat	152400
gttaaatatc	aaagttcagg	ccaggtgcag	tggtcacgc	tagtcttgct	gaacgaaaga	152460
gaggccaag	cgggtgaatg	acttgaggtc	aggagttcaa	ctgtaatccc	agcactttgg	152520
tgaaccctg	tcttagccgg	gtgtggtggc	agggacctgt	gaccagcctg	gccaacatgg	152580
ctgaggcagg	agaatcgctt	gaaccagggg	ggcggaggtt	aatcccaact	atttgggagg	152640
				gcagtgagcc	gagatcacac	152700

104/122

cactgcactc	cagcctgggc	gacgagcgaa	accccatctc	aaaaaaaaaa	tcaaagttca	152760
gagagctcaa	tttgagtaga	agttgtagga	taaggtagca	gaaaagagga	agctgcccag	152820
aaagaaagcc	gtagagatat	ttagagagat	tcccatggat	ccttggccta	ggagtgatct	152880
gtatatgtgt	gggggtgaaaa	cgcatgtgtc	caggtagaga	acccccaga	aattagtagg	152940
ctgaatgatt	gctggaacat	agggctaaga	aaagttcatg	gccagaagga	tctggccaga	153000
gtagagagac	ttagtaatac	acaaggcatt	gggtagtgtc	ttcacagagg	ttatgcctta	153060
ctactgaaga	taaattagtc	ctagagtaca	agcacctgaa	ccaagtttca	aagcaaattt	153120
ttaaagggtc	aaattaccta	acaactgcat	gccaaaacaa	aggcctaacc	ctctttacag	153180
taacacaaca	aaattcagca	cttcacagtg	taaagttaga	atgtctgacg	tccaggctgg	153240
gcgcagtgcc	tcatgcctgt	aatcccagca	ctttgggagg	ccgaggcagg	tagatgacct	153300
gaggtcagga	gttcaagacc	agcctggcta	acatgggtga	accccgcttc	tattaaaaat	153360
acaaaaactt	agccaggcat	ggtagccggc	acctgtgac	ccggctaact	ggaggctga	153420
ggcaggagaa	ttgcctgaac	ccaggaggtg	aaggttgacg	tgagccgaga	tcgacacct	153480
gcactctggt	ctggggcaaaa	agagcaaaac	tcaggctcaa	aaaaaaaaaa	gaatgtctga	153540
cgtcaatcac	aaattaccaa	gcatgacatg	aagttgacct	ataaccagga	gaaaactcaa	153600
tctatagaaa	cagacccaga	tgtgagaaat	atgtgaaatt	tagcagacaa	agaccatcaa	153660
gtggctatct	taaataattaa	aaatatgttc	aagtgccag	gtgcagtggc	tcatgcctgt	153720
aatcccagca	ctttgggagg	ccaaggtggg	taggagttca	agaccagctt	ggccaatatg	153780
gtgaaacccc	ttctctacta	aaaatacaaa	aaaattagct	gggcatgggtg	gcagggtgcct	153840
atagtcccag	ctatatggga	ggctgaggca	caagaatcac	ttgaaccccg	gaggtggagg	153900
ttgaggttgc	agtaagccga	gattgtgcca	cttgacttcc	agcctggaca	acagagttag	153960
actctgtctc	aaaaaaaaaa	aaaaaaaaagt	taaagaaaac	aagagtataa	tgagaaaaat	154020
gcaaaatagt	tttaaaagaa	ccaaatggaa	tttcttaaaa	taaaaaatac	cagaaatggg	154080
ggccggggcg	ggtagctcac	gtctataatc	ccagcacttt	gtgggggctg	aggcaggctag	154140
atcacctgag	atcggtagtt	caaggccagc	tggaaccaata	tggagaaaacc	tcactctctac	154200
taaaaaatata	aaattagctg	ggcgtgggtg	cgcattgcct	gtaatcccag	ctacttggga	154260
ggctgaggca	ggagaattgc	ttgaaccccg	gagggcagagg	ttgcgggtgag	ctgagattgc	154320
accagtgcac	tccagcttgg	gccacaagag	tgaaactccg	tctcaaaaaa	aaaacaaaaa	154380
aaaacagtag	actcgaagaa	ctagctgagt	tttctactac	tttaggcagt	aagtgtgacc	154440
ttttgcaggt	gactacttta	gttcctcatg	tcctcattag	tagatcagag	aaattcgaca	154500
ccaaaacccc	aaaagaaaaa	ccccttctaa	tcctcattcc	atgattttat	gaatgcatga	154560
agtcctaggg	ctgcgaagga	atactcatte	tctttatcct	gtgttgatac	ctctctgctt	154620
caacctccaa	gcttatagga	gcctatagga	cattcagcat	cattcagcat	aaactacctt	154680
acaccattac	tgaattgctt	catgtgcaca	tgtcccatgc	cacaataaccg	gggaccttgt	154740
cttcctgtat	atgtgtccgc	agtgtctgtg	ctacaggagg	gagtcagtga	atgtctgcat	154800
gtgtgtcttt	accatccctc	ttgaatatgc	tctagggtta	attcctagaa	gtagaattac	154860
tctattgaaa	attgggaata	tttttctatt	taatatctat	tgccaacatg	ggaaagcaag	154920
tctggatgcc	agtccttggt	atatgcccct	tgggtaagtt	acgtaacctc	tttaagcttc	154980
tggtcactca	tattttaaca	aggaaaatta	caatatttta	cctcacaaaa	ttgtagtcag	155040
cttctgggtg	tcttaaaact	tggatatatg	taaacactaa	gtgttggtgt	ccatccttaa	155100
tttgtaataa	taggtcactt	gttagagaaa	tgcaccttac	cattttcttt	tcttttcttt	155160
tttcagttat	gactcaaaac	ttgagataaa	ggaaatctgc	ttgtgaaaaa	taagagaact	155220
tttttccctt	gggtggattc	ttcaacacag	ccaatgaaaa	cagcactata	tttctgatct	155280
gtcactgttg	tttccaggag	agaatgggag	acaatcctag	acttccacca	taatgcagtt	155340
acctgtaggc	ataattgatg	cacatgatgt	tcacacagtg	agagtcttaa	agatacaaaa	155400
tggtattggt	tacattacta	gaaaattatt	agttttccaa	tggcaataac	ccatttatga	155460
gagtggttta	gcctactgga	atagacaggg	accacatcct	ctgggaagca	gataagcata	155520
gaactgatac	ttgatgcaca	ctcgtagtgg	taactcatcc	ctaatacagca	ttgtaaagca	155580
gggtgccagag	gtggtttgct	ttgtccttcc	aaagcagggtg	agtcagcccc	accgagagcc	155640
aggcagcttt	gagtggcagc	gtggtgctag	cagcttcagc	ggaacagggt	gagagttaat	155700
tatgcagtct	tcttgacagc	ggcatttaatt	tggaaggaaa	ctgacaagtc	atgggtcaag	155760
tttcagtgac	ttcctccttc	ctctgatggc	agtatatagt	tttcacattt	taattcctcc	155820
tctgagatg	cactatactt	aaaaccattc	tctccccctgc	taacagaagg	gtgtgaatct	155880
gggtttacttt	gagcattagg	atgtgcccct	ttggaaattct	gcactccagt	tacttaactt	155940
tcccttcaga	atacatgtgg	aaagaaagaa	agaaatagcg	atgactccac	ttttgcccct	156000
gtggcacctt	gaacaaagca	gttcttccca	aattatactt	tttttttttt	taaaataagg	156060
gagcaggatg	actggggaga	gagaaacatt	tgactttgac	tgctcccccc	attcctttgt	156120
gtgagctgga	aagtgtgcag	ttggtcgtct	ttcttttagga	tgctcccccc	tagtaagaga	156180
ctcactcact	gcacttctgc	tcagttaggt	tctgcatcgg	gatcacacag	ccatcagcag	156240
gactgcccag	ttggtgagca	cactccattg	accacgcggc	gccagcgctt	cctcaatgca	156300

105/122

catgattgag	aggaaagaaa	gttctcttag	atgttactgc	ttttgctcag	acttttgcaaa	156360
aaaaaaaaata	tatatatata	tgtataaata	tataattatt	aatcactttt	gtccttgaga	156420
aagtcctgaa	tgaacagaga	at ttattcca	ttgcaatatt	tgattgtata	gaggcacact	156480
gtttcaatcga	cagaagaagc	aaaaaggctt	tgtgtaaagt	tttgggtacta	tgtaccacct	156540
ctgttattct	tttaaaagctg	aagtattcat	gtacttaaac	catattatat	ttaattgtgt	156600
ttgatttttaa	aatatatata	tatgaattct	at tttaaaatt	gtgtcaactt	tctgctttca	156660
gggcatttat	ggctcttctg	ttgaaatata	ttgatctttc	caaataattt	cat ttgcttt	156720
ctaaaaaccc	agaacatgag	ccactactgg	actttgcctt	gtgtttgaag	tgtagggcat	156780
aaacccaagg	tttttattag	tcatctatgc	tgtagttaat	tcattttgtt	cttttaacaa	156840
aatatttcca	tccacttcac	attgcttcaa	tctttaacag	aaaagcaata	taaaggttat	156900
agaataaaaat	gtgggttttg	gcaactcttg	ctgcctctgc	atgttttgga	ataacaattt	156960
ctacaagact	ctaggctggt	taaactagt	ctttcagtta	agataaattc	taatcatttc	157020
tttgtatata	cat ttgtg	ttctgagcta	gagatgcaa	gtagtgttaa	actgcttata	157080
aagagaatag	cagcaaattt	gagactcggc	tacttttttc	tgccccacct	gctttgagac	157140
acagaagcgg	agtgtggccc	gaaattatta	gccagattta	atatttgatc	taaagtaggt	157200
ccttgtactc	at tttaagt	tggaatttga	tctctccaac	attgagcacc	caccatgttc	157260
caggctctgt	gcattgtgcc	cacaaaataa	gattccctgg	tggagttttt	atgggttcaa	157320
ataatcagtt	gaacaccctt	catctttatc	atgtttgttga	cattgacaca	aattgtttta	157380
aaagaaaaaga	tattagagag	aaagtgggtac	ctttgtaaact	tgatgtgtct	tcatcattcg	157440
gtaagatttg	atgaaagtaa	aaagcaaatg	tcagccaaat	ccagtgaaca	gcaataaaac	157500
agggagttaac	tttttataac	tttttctact	tggattttcaa	cattcagtag	agcttttcga	157560
aatgtaagta	gtttacagta	ctggagggtt	gactagttca	gtaggaattt	ggagggggaag	157620
gtcat tctga	attgtaacaa	agtacaaact	tctttgctgt	tttatttaag	tactgagagc	157680
taagcacctg	atggaagtgc	tgacctctct	ccagtgcagc	tgtttgggta	cctgcctgac	157740
ttcaggagtg	gggtttatgt	ttctacacag	tggatttttc	tctcgccctc	tccctcctct	157800
tgccccacaca	ccagttgatt	ggacctgggt	tgaactcctg	atccagacag	gccccagaca	157860
gttcttaatg	ttaagaattt	tggggccggg	cacggtggct	catgcctgta	attgcaacac	157920
tttgggaggc	cgagcagggc	ggatcacttg	aggtcagggg	ttcgaggcca	gcctggccaa	157980
cattgtgaaa	ccctgtcttt	actaaaaata	caaaaattag	ctgggcatgg	tgggcacgc	158040
ctgtaatccc	agctacgtgg	gtggctgaga	caggggaaatc	gcttgaacct	ggaggcggag	158100
gttgtgcaat	gagccgagac	cgtgtcactg	cattccagcc	tgggtgacag	agggagactc	158160
tgtctccaaa	aataaaaaata	agaaaaagaa	ttttgggcta	gggtgcagtgg	ctcacgcctg	158220
taattacagc	at ttggag	gccccagatg	ggcagatcac	ttgaggacag	gagttcgaga	158280
ccagcctgga	caacatgggtg	aaactccatc	tctactaaaa	agacaaaagt	tagccagatg	158340
tgggtgatggg	cacctataat	cctagctcct	cgggaggctg	gggcaggaga	atcacttgaa	158400
cccaggaagc	agagatttga	gtgagccaa	atcacatctc	tgactccag	cctgggcaac	158460
agagcaagac	tctgtctcaa	aaaaaaaaga	at ttggccag	gcgcagtggg	tcacgcctgt	158520
aatcccagca	ctttgggagg	ccaaggcagg	cagatcacga	ggtcaggaga	tcgagattgt	158580
cctggctaac	atgggtgaaac	cctgtctcta	ctaaaaatac	aaaacattag	ccgggtgtgg	158640
tgggtgggcac	ctgtagtccc	agctactagg	gagggtgagg	cagaggaagg	atgtgaaccc	158700
aggaggcgga	gcttgacagta	agccaagatc	gtgccactgc	actacagtct	gggagcagaca	158760
gtgagactcc	gtctcaaaaa	aaaaaagaat	tttggccggg	tgcggtggca	catgcctgta	158820
gtcccagcac	tttgggagac	caaagtgggc	ggattacctg	aggtcaggag	ttcaagacca	158880
gtccggccaa	tatggcgaaa	ccctgtctct	tactaaaaaa	aatacaaaaa	ttagccaggt	158940
gtgggtggcgg	gcacctgggg	aggctgaggc	agggagaaat	gcttgaaccg	gggaggcaga	159000
gggtgcagta	agccaagatc	gtgccactgc	actccagagc	aagactcttt	ctcaaaaaaa	159060
aaaaaaaaaag	aat ttgtcat	ggggaaggag	agatactgtt	caccatctgg	aatgggtgctt	159120
ggatgtggca	cttacaataat	caggagccag	cactgcatgg	acaaacagaa	gcatgtgggc	159180
ctgagatagc	aggtaccttg	ataaccctga	agacatcctt	ggtttctgca	tctattcctg	159240
catccttgca	ttggactaca	ttaatctgtc	agttatcctt	ataatgattt	ttgatttttt	159300
ttttttgaga	tggagtttctg	ctcttgttgc	ccaggctgga	gtgcaatggc	acgatctcgg	159360
ctcaccacaa	cctccacctc	ccagggtcaa	gtgattctgc	tgcttcagcc	tcctgagtaa	159420
ctgggatttac	aggcatgcgc	caccacacct	ggctaatttt	gtatttttag	tagagacggg	159480
gtttctccat	tttggctcagg	ctgggtctga	actcccaacc	tcagggtgatc	accctgtctc	159540
ggcctcccaa	agtgtggga	ttacaggcgt	aagccatggg	accgggtctg	ttttttgatt	159600
ttttgaaacc	agtctgaagt	gagttttttt	aattacgtga	aaggagtttg	gctaaaaatac	159660
tgccatactg	ccctaattgcc	taatgattat	gtattctcga	catgtctgca	aagtactgct	159720
gatttctgga	gaataattttt	tctttagtaa	acttcactta	agtcgtcatg	tgtattctct	159780
caaaatggta	tcctaacctta	atggagctaa	agagaccccc	ttgtttttat	aacaagcagt	159840
tactgaggcc	caggaagggg	agaagtccct	ggcttgtgag	atgatcacca	ttagaactca	159900

106/122

```

ggcctggggcc agtgcctttt catgcttctc agatccttcc aaagaataat gaagattata 159960
accgcttttta gcaattgttaa taaacccaga aatagaaaagc tttttgggta gagtactggg 160020
agaagtttgg cgaggagagat aatttttaca aaatttgtaa atacctgcca attctatata 160080
ctaggcaagg tctctggcct tgtaaaaccc ctcaagggta caactttggg ggcccacact 160140
aatagttacc cactgaggcc ctctccgggt gaacattgag cactagagga agccccctctg 160200
cttggggcagg actggggcgtg gtgcagagta ggagcgggta tactgtggat tctggggcagg 160260
tggagatggc cagtgatgtc caataaagga cactggaggg agcagtgtga gtaaaggccc 160320
tgagggcatt catgttcagg gagggttgct gccactggc ttgcttggca cacaggagag 160380
tgggtattcc tgccttagta actttatgta aacaagtatt tcctcagtct gttccttctca 160440
aactgcctgc tctggcacat tcagaatgtc acagaactca cctggatgca ttcagcccct 160500
tgcctaaagg tgacagtgca tctccttccc caccaccacc ctcataccac tgaagcacct 160560
gtcagactgg ccagctctgt gggcaaggag cctagagagg gcttagtttc agcttgaaaag 160620
gagctgggat ttaccaagaa gcaaatgaga gacgaggatt gcaacaactg tgccatttcc 160680
ccagcttcag ctgactcctg tatattgact gtgccttcag actcatccgt aagtgacccc 160740
aggctggcct ctcccacatc acagtaagaa ttccacacac catacaactt ggaaaggaggc 160800
tccagctgaa ggaagcccca cacttcttcc aagtttttct tagtcttctc ttcttggcaa 160860
agagtacctt ttgtttcttc taattatgta actattgggt tagtaaatat tcacccattc 160920
agtacacctg taagtggcag gcactgttta cagggacaca ggaaggaata aaaacttgca 160980
ggcaccttgg agcttgcat ttttgaaga ggtaatggaa gttgggtag cagctaaact 161040
atgtcggat tgggcaggcg cagtggctca cacctgtaat cccagcactt tggaggccaa 161100
ggtagggcaga tcatgaagtc aggagatcga gaccatcctg gctaaccatgg tgaaaccccg 161160
tctctactaa aagtaaaaaa aaaaattagc caggtgtggt ggcgggcgcc tgtagtccca 161220
gctacttggg aggctgaggc aggagaatgg tgtgaaccca ggaggcggaag attgcagtga 161280
gccgagatgg caccactgca ctccagcctg ggtgacagag cgagactctg tctcagaaaa 161340
aaaaaatatg ctggtagatt tgattcaaga tggcctttgg agcccatgat ttaggtctcg 161400
taccaccaa ggtctactgg aaaacatcag gctctcctgc tatagacca tagggagagc 161460
tgcagccgag aggggggagct gaagagaagt gccccttctg tgcctgtca gcctcatcct 161520
tccgcaagga ccagttgctg tgccactcca ttcacttgct gcaagactgg aggtttttcc 161580
tcagggtgtg agcacctggg ttacaagatg tcagcatcct gatgcctgag accatcaagg 161640
caagtctctg aacagggctt accttagagt aaggcttaga agaggccgta aagtcagtct 161700
cagctccgtg gctctgcaga gctttgggac atgtgaattc ttaaaaacaa gactattgta 161760
cagttactat atgcatgcag tataaaatta taaccttgga aaatcctagc tagctgttga 161820
gctaaatcca taaagtaatt agctcctgag ttctgaggt gtaataataa tcagcataat 161880
gagtaaacac tgtgtgtgcc aggcagcgtc tcatattgat cttgtgataa tcttgtaagt 161940
actgattttc tcccttcttt aaacaaagtt tttttttttt ttttagagag ggtctcacta 162000
tgttgcccag gctagtcttg aattc 162025

```

<210> 37  
<211> 1350  
<212> DNA  
<213> Homo Sapien

<220>  
<221> CDS  
<222> (213) ... (920)

<300>  
<308> GenBank AJ242973  
<309> 1999-10-26

```

<400> 37
gcgggcgcgt cgacgtgaca gccgggtacgc ccgggttttg gcaacctcga ttacggggcgg 60
cctccaggcc cgccagcagc gccccgcgcc gcccgcgccg gcccctgccg cccccgggtt 120
ccggcgcgccg accccactct ctgccgttcc ccgtgcgggt tagcgccgtc 180
ccccgggacc acccttcggc tggcgccctc cc atg ctc tgc gcc acc cgg agg 233
Met Leu Ser Ala Thr Arg Arg
1 5

```

```

gct tgc cag ctc ctc ctc ctc cac agc ctc ttt ccc gtc ccg agg atg 281
Ala Cys Gln Leu Leu Leu Leu His Ser Leu Phe Pro Val Pro Arg Met

```

107/122

10	15	20	
ggc aac tcg gcc tcg aac atc gtc agc ccc cag gag gcc ttg ccg ggc Gly Asn Ser Ala Ser Asn Ile Val Ser Pro Gln Glu Ala Leu Pro Gly 25 30 35			329
cgg aag gaa cag acc cct gta gcg gcc aaa cat cat gtc aat ggc aac Arg Lys Glu Gln Thr Pro Val Ala Ala Lys His His Val Asn Gly Asn 40 45 50 55			377
aga aca gtc gaa cct ttc cca gag gga aca cag atg gct gta ttt gga Arg Thr Val Glu Pro Phe Pro Glu Gly Thr Gln Met Ala Val Phe Gly 60 65 70			425
atg gga tgt ttc tgg gga gct gaa agg aaa ttc tgg gtc ttg aaa gga Met Gly Cys Phe Trp Gly Ala Glu Arg Lys Phe Trp Val Leu Lys Gly 75 80 85			473
gtg tat tca act caa gtt ggt ttt gca gga ggc tat act tca aat cct Val Tyr Ser Thr Gln Val Gly Phe Ala Gly Gly Tyr Thr Ser Asn Pro 90 95 100			521
act tat aaa gaa gtc tgc tca gaa aaa act ggc cat gca gaa gtc gtc Thr Tyr Lys Glu Val Cys Ser Glu Lys Thr Gly His Ala Glu Val Val 105 110 115			569
cga gtg gtg tac cag cca gaa cac atg agt ttt gag gaa ctg ctc aag Arg Val Val Tyr Gln Pro Glu His Met Ser Phe Glu Glu Leu Leu Lys 120 125 130 135			617
gtc ttc tgg gag aat cac gac ccg acc caa ggt atg cgc cag ggg aac Val Phe Trp Glu Asn His Asp Pro Thr Gln Gly Met Arg Gln Gly Asn 140 145 150			665
gac cat ggc act cag tac cgc tcg gcc atc tac ccg acc tct gcc aag Asp His Gly Thr Gln Tyr Arg Ser Ala Ile Tyr Pro Thr Ser Ala Lys 155 160 165			713
caa atg gag gca gcc ctg agc tcc aaa gag aac tac caa aag gtt ctt Gln Met Glu Ala Ala Leu Ser Ser Lys Glu Asn Tyr Gln Lys Val Leu 170 175 180			761
tca gag cac ggc ttc ggc ccc atc act acc gac atc cgg gag gga cag Ser Glu His Gly Phe Gly Pro Ile Thr Thr Asp Ile Arg Glu Gly Gln 185 190 195			809
act ttc tac tat gcg gaa gac tac cac cag cag tac ctg agc aag aac Thr Phe Tyr Tyr Ala Glu Asp Tyr His Gln Gln Tyr Leu Ser Lys Asn 200 205 210 215			857
ccc aat ggc tac tgc ggc ctt ggg ggc acc ggc gtg tcc tgc cca gtg Pro Asn Gly Tyr Cys Gly Leu Gly Gly Thr Gly Val Ser Cys Pro Val 220 225 230			905
ggt att aaa aaa taa ttgctcccca catggtgggc ctttgaggtt ccagtaaaaa Gly Ile Lys Lys * 235			960
tgctttcaac aaattgggca atgcttgtgt gattcacaat cgtggcattt aaagtgcaca aagtacaaag gaatttatac agattggggtt taccgaagta taatctatag gaggcgcgat			1020 1080

108/122

```

ggcaagttga taaaatgtga cttatctcct aataagttat ggtgggagtg gagctgtgcg 1140
gtttcctgtg tcttctgggg tctgagtga gatagcaggg atgctgtgtt cacccttctt 1200
ggtagaagct aaggtgtgag ctgggaggtt gctggacagg atgggggacc ccagaagtcc 1260
tttatctgtg ctctctgccc gccagtgcct tacaatttgc aaacgtgtat agcctcagtg 1320
actcattcgc tgaaatcctt cgctttacca 1350

```

<210> 38  
<211> 235  
<212> PRT  
<213> Homo Sapien

<400> 38

Met	Leu	Ser	Ala	Thr	Arg	Arg	Ala	Cys	Gln	Leu	Leu	Leu	Leu	His	Ser
1				5					10					15	
Leu	Phe	Pro	Val	Pro	Arg	Met	Gly	Asn	Ser	Ala	Ser	Asn	Ile	Val	Ser
			20					25					30		
Pro	Gln	Glu	Ala	Leu	Pro	Gly	Arg	Lys	Glu	Gln	Thr	Pro	Val	Ala	Ala
			35				40					45			
Lys	His	His	Val	Asn	Gly	Asn	Arg	Thr	Val	Glu	Pro	Phe	Pro	Glu	Gly
	50				55					60					
Thr	Gln	Met	Ala	Val	Phe	Gly	Met	Gly	Cys	Phe	Trp	Gly	Ala	Glu	Arg
65				70					75					80	
Lys	Phe	Trp	Val	Leu	Lys	Gly	Val	Tyr	Ser	Thr	Gln	Val	Gly	Phe	Ala
			85					90					95		
Gly	Gly	Tyr	Thr	Ser	Asn	Pro	Thr	Tyr	Lys	Glu	Val	Cys	Ser	Glu	Lys
			100					105					110		
Thr	Gly	His	Ala	Glu	Val	Val	Arg	Val	Val	Tyr	Gln	Pro	Glu	His	Met
		115					120					125			
Ser	Phe	Glu	Glu	Leu	Leu	Lys	Val	Phe	Trp	Glu	Asn	His	Asp	Pro	Thr
		130				135					140				
Gln	Gly	Met	Arg	Gln	Gly	Asn	Asp	His	Gly	Thr	Gln	Tyr	Arg	Ser	Ala
145				150					155					160	
Ile	Tyr	Pro	Thr	Ser	Ala	Lys	Gln	Met	Glu	Ala	Ala	Leu	Ser	Ser	Lys
			165					170					175		
Glu	Asn	Tyr	Gln	Lys	Val	Leu	Ser	Glu	His	Gly	Phe	Gly	Pro	Ile	Thr
		180						185					190		
Thr	Asp	Ile	Arg	Glu	Gly	Gln	Thr	Phe	Tyr	Tyr	Ala	Glu	Asp	Tyr	His
		195				200						205			
Gln	Gln	Tyr	Leu	Ser	Lys	Asn	Pro	Asn	Gly	Tyr	Cys	Gly	Leu	Gly	Gly
	210				215						220				
Thr	Gly	Val	Ser	Cys	Pro	Val	Gly	Ile	Lys	Lys					
225					230					235					

<210> 39  
<211> 481  
<212> DNA  
<213> Homo Sapien

<300>  
<308> GenBank AW195104  
<309> 1999-11-29

<400> 39

ggcattattg	gactgtaggt	ttttattaaa	acaaacattt	ctcatagctc	taagcaaagc	60
attagaattc	atcaagcgga	ctcacatctt	ttctctgcac	agagaggggc	tgaaaagggg	120
gagaaagtcc	cttatgtatg	tctagatttg	gtaaagcgaa	ggatttcagc	gaatgagtca	180
ctgaggctat	acacgtttgc	aaattgtaag	gcactggcgg	gcagagagca	cagataaagg	240
acttctgggg	tcccccatcc	tgtccagcaa	cctcccagct	cacaccttag	cttctaccaa	300
gaaggggtgaa	cacagcatcc	ctgctatctt	cactcagacc	ccagaaaacc	cagggaaacc	360
cgacagctcc	actcccacca	taacttatta	ggagataagt	cacattttat	caacttgcca	420

109/122

tcgcgcctcc tatagattat acttcggtaa acccaatctg tataaattcc tttgtacttt 480  
g 481

<210> 40  
<211> 390  
<212> DNA  
<213> Homo Sapien

<300>  
<308> GenBank AW874187  
<309> 2000-05-22

<400> 40  
ttttttttat tggactgtag gtttttatta aaacaaacat ttctcatagc tctaagcaaa 60  
gcattagaat tcatcaagcg gactcacatc ttttctctgc acagagaggg ctgaaaaggg 120  
agagaaaagg ccttatgtat gtctagattt ggtaaagcga aggatttcag cgaatgagtc 180  
actgaggcta tacacgtttg caaattgtaa ggcactggcg ggcagagagc acagataaag 240  
gacttttggg ggtcccccatt tctgtccag caacctccca gctcacacct tagcttctac 300  
caagaagggg tgaacacagc atccctgcta tcttctactca gacccccaga agacacagga 360  
aaccgcacag ctccactccc accataactt 390

<210> 41  
<211> 43  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 41  
agcggataac aatttcacac agggagctag cttggaagat tgc 43

<210> 42  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 42  
gtccaatata tgcaaacagt tg 22

<210> 43  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 43  
agcggataac aatttcacac agg 23

<210> 44  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>

110/122

<223> Oligonucleotide Primer

<400> 44  
actgagcctg ctgcataa 18

<210> 45  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 45  
tctcaatcat gtgcattgag g 21

<210> 46  
<211> 43  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 46  
agcggataac aatttcacac agggatcaca cagccatcag cag 43

<210> 47  
<211> 23  
<212> DNA  
<213> oligonucleotide primer

<400> 47  
agcggataac aatttcacac agg 23

<210> 48  
<211> 18  
<212> DNA  
<213> Oligonucleotide primer

<400> 48  
ctggcgccac gtggtcaa 18

<210> 49  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 49  
tttctctgca cagagagggc 20

<210> 50  
<211> 44  
<212> DNA  
<213> Artificial Sequence

<220>

111/122

<223> Oligonucleotide Primer

<400> 50  
agcggataac aatttcacac agggctgaaa tccttcgctt tacc 44

<210> 51  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 51  
agcggataac aatttcacac agg 23

<210> 52  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 52  
ctgaaaaggg agagaaag 18

<210> 53  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 53  
tcccaaagtg ctggaattac 20

<210> 54  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 54  
gtccaatata tgcaaacagt tg 22

<210> 55  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 55  
cccacagcag ttaatccttc 20

112/122

<210> 56  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 56  
gcgctcctgt cggtgccca

18

<210> 57  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 57  
gcctgactgg tggggccc

18

<210> 58  
<211> 15  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 58  
catgcatgca cggtc

15

<210> 59  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 59  
cagagagtac ccctcgaccg tgcatgcatg

30

<210> 60  
<211> 15  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 60  
catgcatgca cggtt

15

<210> 61  
<211> 30  
<212> DNA  
<213> Artificial Sequence

113/122

<220>  
<223> Oligonucleotide primer  
  
<400> 61  
gtacgtacgt gccaaactccc catgagagac 30  
  
<210> 62  
<211> 14  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Oligonucleotide primer  
  
<400> 62  
catgcatgca cggt 14  
  
<210> 63  
<211> 18  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Oligonucleotide primer  
  
<400> 63  
gcctgactgg tggggccc 18  
  
<210> 64  
<211> 26  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Oligonucleotide primer  
  
<400> 64  
gtgctgcagg tgtaaacttg taccag 26  
  
<210> 65  
<211> 28  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Oligonucleotide primer  
  
<400> 65  
cacggatccg gtagcagcgg tagagttg 28  
  
<210> 66  
<211> 19  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Oligonucleotide primer  
  
<400> 66  
actgggcatg tggagacag 19

114/122

<210> 67  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 67  
gcactttctt gccatgag

18

<210> 68  
<211> 14  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 68  
tcagtcacga cggt

14

<210> 69  
<211> 14  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 69  
cggataacaa ttct

14

<210> 70  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 70  
caatttcattc gctggatgca atctgggcta tgagatc

37

<210> 71  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 71  
caatttcaca cagcggatgc ttcttttggc tctgact

37

<210> 72  
<211> 40  
<212> DNA  
<213> Artificial Sequence

115/122

<220>  
<223> Oligonucleotide primer

<400> 72  
tcagtcacga cgttggatgc caataaaagt gactctcagc 40

<210> 73  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 73  
cggataacaa tttcggatgc actgggagca ttgaggc 37

<210> 74  
<211> 38  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 74  
tcagtcacga cgttggatga gcagatccct ggacaggc 38

<210> 75  
<211> 38  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 75  
cggataacaa tttcggatgg acaaaatacc tgtattcc 38

<210> 76  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 76  
tcagtcacga cgttggatgc agagcagctc cgagtc 36

<210> 77  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 77  
cagcggtgat cattggatgc aggaagctct gg 32

116/122

<210> 78  
<211> 38  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 78  
tcagtcacga cgttggatgc ccacatgccca cccactac

38

<210> 79  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 79  
cggataacaa tttcggatgc ccgtcaggta ccacg

35

<210> 80  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 80  
tcagtcacga cgttggatgc ccacagtggg gcttcag

37

<210> 81  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 81  
gctcatacct tgcaggatga cg

22

<210> 82  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 82  
tcagtcacga cgttggatga ccagctgttc gtgttc

36

<210> 83  
<211> 34  
<212> DNA  
<213> Artificial Sequence

117/122

<220>  
<223> Oligonucleotide primer

<400> 83  
tacatggagt tcggggatgc acacggcgac tctc 34

<210> 84  
<211> 40  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 84  
tcagtcacga cggttgatgg ggaagagcag agatatacgt 40

<210> 85  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 85  
gaggggctga tccaggatgg gtgctccac 29

<210> 86  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 86  
tgaagcactt gaaggatgag ggtgtctgcg 30

<210> 87  
<211> 38  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 87  
cggataacaa ttctggatgc tgcgtgatga tgaaatcg 38

<210> 88  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 88  
gatgaagctc ccaggatgcc agaggc 26

118/122

<210> 89  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 89  
gccgccggtg taggatgctg ctggtgc

27

<210> 90  
<211> 31  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Template

<400> 90  
cgcagggttt cctcgtcgca ctgggcatgt g

31

<210> 91  
<211> 43  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Biotinylatd primer

<400> 91  
tgcttatccc ttagctacc ctgtcttggc cttgcagatc caa

43

<210> 92  
<211> 42  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 92  
agcggataac aatttcacac aggccatcac accgcggtac tg

42

<210> 93  
<211> 44  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide primer

<400> 93  
cccagtcacg acgttgtaaa acgtcttggc cttgcagatc caag

44

<210> 94  
<211> 42  
<212> DNA  
<213> Artificial Sequence

119/122

<220>  
 <223> Oligonucleotide primer  
 <400> 94  
 agcggataac aatttcacac aggccatcac accgcggtac tg 42  
 <210> 95  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Oligonucleotide primer  
 <400> 95  
 ctccagctgg gcaggagtgc 20  
 <210> 96  
 <211> 17  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Oligonucleotide primer  
 <400> 96  
 cacttcagtc gctccct 17  
 <210> 97  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Biotinylated primer  
 <400> 97  
 cccagtcacg acgttgtaaa acg 23  
 <210> 98  
 <211> 100  
 <212> DNA  
 <213> Homo sapien  
 <400> 98  
 cctttgagaa agggctctgc ttgagttgta gaaagaaccg ctgcaacaat ctgggctatg 60  
 agatcaataa agtcagagcc aaaagaagca gcaaaatgta 100  
 <210> 99  
 <211> 100  
 <212> DNA  
 <213> Homo sapien  
 <400> 99  
 cctttgagaa agggctctgc ttgagttgta gaaagaaccg ctgcaacaat ctgggctatg 60  
 agatcagtaa agtcagagcc aaaagaagca gcaaaatgta 100  
 <210> 100  
 <211> 100  
 <212> DNA

120/122

&lt;213&gt; Homo sapien

&lt;400&gt; 100

```
gaattatattt tgtgttttcta aaactatggg tcccaataaa agtgactctc agcgagcctc    60
aatgctccca gtgctattca tgggcagctc tctgggctca    100
```

&lt;210&gt; 101

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 101

```
gaattatattt tgtgttttcta aaactatggg tcccaataaa agtgactctc agcaagcctc    60
aatgctccca gtgctattca tgggcagctc tctgggctca    100
```

&lt;210&gt; 102

&lt;211&gt; 84

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 102

```
taataggact acttctaate tgtaagagca gatccctgga caggcgagga atacaggtat    60
ttgtccttg aagtaacctt tcag    84
```

&lt;210&gt; 103

&lt;211&gt; 84

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 103

```
taataggact acttctaate tgtaagagca gatccctgga caggcaagga atacaggtat    60
ttgtccttg aagtaacctt tcag    84
```

&lt;210&gt; 104

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 104

```
ctcaccatgg gcatttgatt gcagagcagc tccgagtccg tccagagctt cctgcagtca    60
atgatcaccg ctgtgggcat ccctgaggtc atgtctcgta    100
```

&lt;210&gt; 105

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 105

```
ctcaccatgg gcatttgatt gcagagcagc tccgagtcca tccagagctt cctgcagtca    60
atgatcaccg ctgtgggcat ccctgaggtc atgtctcgta    100
```

&lt;210&gt; 106

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 106

```
agcaaggact cctgcaaggg ggacagtggg ggccacatg ccaccacta ccagggcacg    60
tggtagctga cgggcatcgt cagctggggc cagggctgcg    100
```

121/122

&lt;210&gt; 107

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 107

agcaaggact cctgcaaggg ggacagtgga ggcccacatg ccaccacta ccggggcacg 60  
tggtacctga cgggcatcgt cagctggggc cagggctgcg 100

&lt;210&gt; 108

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Hom sapien

&lt;400&gt; 108

caataactct aatgcagcgg aagatgacct gccacagtg gagcttcagg gcgtggtgcc 60  
ccggggcgctc aacctgcaag gtatgagcat accccccttc 100

&lt;210&gt; 109

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 109

caataactct aatgcagcgg aagatgacct gccacagtg gagcttcagg gcttggtgcc 60  
ccggggcgctc aacctgcaag gtatgagcat accccccttc 100

&lt;210&gt; 110

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 110

ttgaagcttt gggctacgtg gatgaccagc tgttcgtgtt ctatgatcat gagagtcgcc 60  
gtgtggagcc ccgaactcca tgggtttcca gtagaatttc 100

&lt;210&gt; 111

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 111

ttgaagcttt gggctacgtg gatgaccagc tgttcgtgtt ctatgatgat gagagtcgcc 60  
gtgtggagcc ccgaactcca tgggtttcca gtagaatttc 100

&lt;210&gt; 112

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 112

ggataacctt ggctgtaccc cctggggaag agcagagata tacgtgccag gtggagcacc 60  
caggcctgga tcagcccctc atttgtatct gggagccctc 100

&lt;210&gt; 113

&lt;211&gt; 100

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 113

122/122

ggataacctt ggctgtaccc cctggggaag agcagagata tacgtaccag gtggagcacc 60  
 caggcctgga tcagcccctc attgtgatct gggagccctc 100

<210> 114  
 <211> 80  
 <212> DNA  
 <213> Homo sapien

<400> 114  
 tgaagcactt gaaggagaag gtgtctgcgg gagccgattt catcatcacg cagcttttct 60  
 ttgaggctga cacattcttc 80

<210> 115  
 <211> 80  
 <212> DNA  
 <213> Homo sapien

<400> 115  
 tgaagcactt gaaggagaag gtgtctgcgg gagtcgattt catcatcacg cagcttttct 60  
 ttgaggctga cacattcttc 80

<210> 116  
 <211> 80  
 <212> DNA  
 <213> Homo sapien

<400> 116  
 tccagatgaa gctcccagaa tgccagaggc tgctccccgc gtggcccctg caccagcagc 60  
 tcctacaccg gcggcccctg 80

<210> 117  
 <211> 80  
 <212> DNA  
 <213> Homo sapien

<400> 117  
 tccagatgaa gctcccagaa tgccagaggc tgctccccgc gtggcccctg caccagcagc 60  
 tcctacaccg gcggcccctg 80

<210> 118  
 <211> 48  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Hair pin structure

<400> 118  
 cagagagtac ccctcaaccg tgcattgcatt aaacatgcatt gcacgggtt 48

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 April 2001 (19.04.2001)

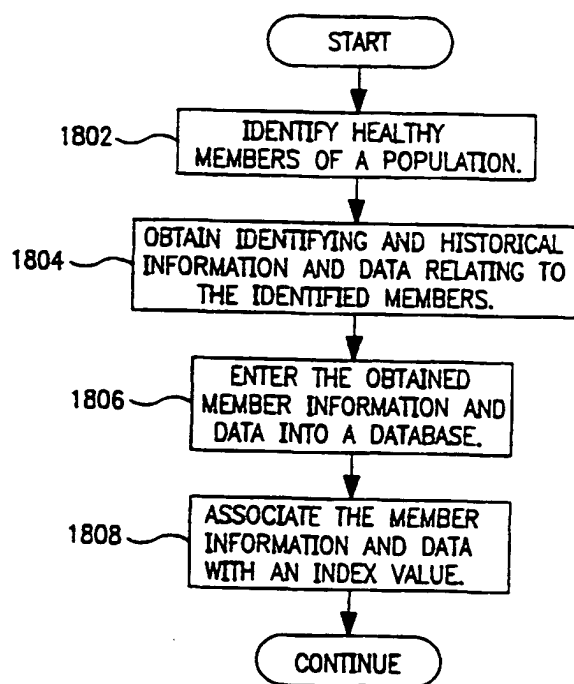
PCT

(10) International Publication Number  
**WO 01/027857 A3**

- (51) International Patent Classification<sup>7</sup>: **G06F 19/00** US 09/663,968 (CIP)  
Filed on 19 September 2000 (19.09.2000)  
US 60/217,251 (CIP)  
Filed on 10 July 2000 (10.07.2000)
- (21) International Application Number: PCT/US00/28413
- (22) International Filing Date: 13 October 2000 (13.10.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/159,176 13 October 1999 (13.10.1999) US  
60/217,658 10 July 2000 (10.07.2000) US  
60/217,251 10 July 2000 (10.07.2000) US  
09/663,968 19 September 2000 (19.09.2000) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:  
US 60/159,176 (CIP)  
Filed on 13 October 1999 (13.10.1999)  
US 60/217,658 (CIP)  
Filed on 10 July 2000 (10.07.2000)
- (71) Applicant (for all designated States except US): **SEQUENOM, INC.** [US/US]: 3595 John Hopkins Court, San Diego, CA 92121 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **BRAUN, Andreas** [DE/US]: 11237-6 Carmel Creek Road, San Diego, CA 92130 (US). **KÖSTER, Hubert** [DE/CH]: Via Delle Scuole 1, CH-6900 Lugano-Cassarate (CH). **VAN DEN BOOM, Dirk** [DE/DE]: Eppendorfer Weg 205 D, D-20253 Hamburg (DE). **PING, Yip** [US/US]: 3641 Copley Avenue, San Diego, CA 92116 (US). **RODI, Charlie** [US/US]: 13823 Recuerdo Drive, Del Mar, CA 92014 (US). **HE, Liyan** [CN/US]: 10948 Creek Bridge Place, San Diego, CA 92128 (US). **CHIU, Norman** [CA/US]: 1128 Caminito Alvarez, San Diego, CA 92126 (US). **JURINKE, Christian** [DE/DE]: Rombergstrasse 22, 20255 Hamburg (DE).

[Continued on next page]

(54) Title: METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING POLYMORPHIC GENETIC MARKERS



(57) Abstract: Process and methods for creating a database of genomic samples from healthy human donors. methods that use the database to identify and correlate polymorphic genetic markers and other markers with diseases and conditions are provided.

WO 01/027857 A3



(74) Agents: SEIDMAN, Stephanie, L. et al.: Heller Ehrman White & McAuliffe, Suite 700, 4250 Executive Square, La Jolla, CA 92037 (US).

patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(81) Designated States (national): AE, AG, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LI, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

**Published:**

— with international search report

(88) Date of publication of the international search report:

3 October 2002

(84) Designated States (regional): ARIPO patent (GH, GM, KI, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# INTERNATIONAL SEARCH REPORT

Intern Application No  
PCT/US 00/28413

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 G06F19/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, EPO-Internal

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 05323 A (AFFYMETRIX INC) 4 February 1999 (1999-02-04) abstract; claims 1,4	1-100
X	WO 97 40462 A (SPECTRA BIOMEDICAL INC) 30 October 1997 (1997-10-30) page 4, line 2 - line 22 page 7, line 3 -page 8, line 4	1-100
X	WO 98 24935 A (AN GANG ;HARA MARK O (US); RALPH DAVID (US); VELTRI ROBERT (US); U) 11 June 1998 (1998-06-11) page 4, line 27 -page 5, line 6 page 6, line 14 - line 18	1-100
	-/-	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*G\* document member of the same patent family

Date of the actual completion of the international search

10 September 2001

Date of mailing of the international search report

28/09/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Fillooy García, E

# INTERNATIONAL SEARCH REPORT

Intern. # Application No  
PCT/US 00/28413

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	COLLINS F S ET AL: "A DNA Polymorphism Discovery Resource for Research on Human Genetic Variation" GENOME RESEARCH, vol. 8, 1998, pages 1229-1231, XP002177106 the whole document	1-100
P, X	WO 00 51053 A (BRYANT STEPHEN PAUL ; GEMINI RESEARCH LTD (GB); KELLY PAUL JAMES (G) 31 August 2000 (2000-08-31) abstract; claims 1-24	1-100
Y	WO 98 35609 A (HELMS RONALD W ; TOMASKO LISA (US); BIOMAR INTERNATIONAL INC (US);) 20 August 1998 (1998-08-20) abstract; claims 1-25	1-100
Y	SARKAR C ET AL: "Human Genetic Bi-allelic Sequences (HGBASE), a Database of Intra-genic Polymorphisms" MEM INST OSWALDO CRUZ, 'Online! vol. 93, no. 5, September 1998 (1998-09) - October 1998 (1998-10), pages 693-694, XP002177107 Rio de Janeiro Retrieved from the Internet: <URL:http://www.cgb.ki.se/cgb/groups/brookes/publications.htm> 'retrieved on 2001-09-05! the whole document	1-100
A	FOSTER M W AND FREEMAN W L: "Naming Names in Human Genetic Variation Research" GENOME RESEARCH, vol. 8, 1998, pages 755-757, XP002177108 the whole document	2,9,27, 55,72

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/28413

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9905323 A	04-02-1999	EP 1002264 A	24-05-2000
		EP 1009861 A	21-06-2000
		EP 0998697 A	10-05-2000
		EP 1007737 A	14-06-2000
		WO 9905574 A	04-02-1999
		WO 9905324 A	04-02-1999
		WO 9905591 A	04-02-1999
		US 6229911 B	08-05-2001
		US 6188783 B	13-02-2001
WO 9740462 A	30-10-1997	AU 2734197 A	12-11-1997
		EP 0897567 A	24-02-1999
		JP 2000508912 T	18-07-2000
WO 9824935 A	11-06-1998	AU 722819 B	10-08-2000
		AU 5515198 A	29-06-1998
		EP 0960214 A	01-12-1999
		US 6190857 B	20-02-2001
WO 0051053 A	31-08-2000	AU 2815900 A	14-09-2000
WO 9835609 A	20-08-1998	US 6059724 A	09-05-2000
		AU 6151498 A	08-09-1998
		BR 9807366 A	18-04-2000
		CN 1268033 T	27-09-2000
		EP 0973435 A	26-01-2000

CORRECTED VERSION

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 April 2001 (19.04.2001)

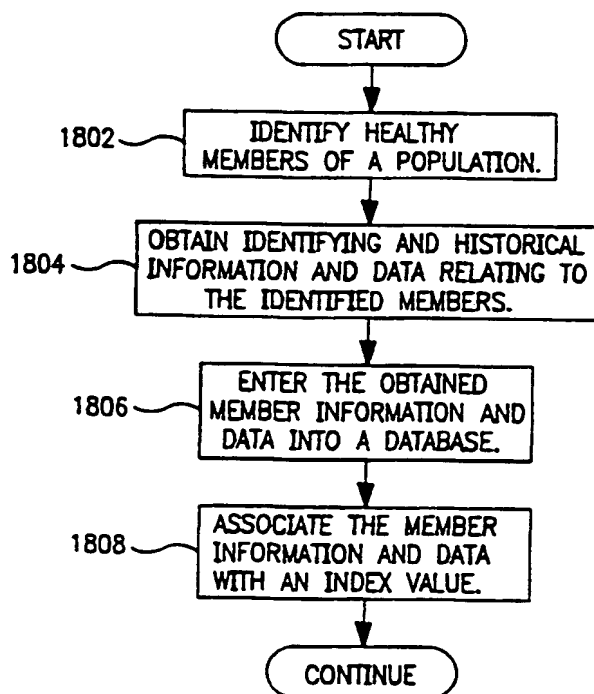
PCT

(10) International Publication Number  
**WO 01/027857 A3**

- (51) International Patent Classification<sup>7</sup>: **G06F 19/00** US 60/217,658 (CIP)  
Filed on 10 July 2000 (10.07.2000)  
(21) International Application Number: PCT/US00/28413 US 09/663,968 (CIP)  
Filed on 19 September 2000 (19.09.2000)  
(22) International Filing Date: 13 October 2000 (13.10.2000) US 60/217,251 (CIP)  
Filed on 10 July 2000 (10.07.2000)  
(25) Filing Language: English  
(26) Publication Language: English  
(30) Priority Data:  
60/159,176 13 October 1999 (13.10.1999) US  
60/217,658 10 July 2000 (10.07.2000) US  
60/217,251 10 July 2000 (10.07.2000) US  
09/663,968 19 September 2000 (19.09.2000) US  
(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:  
US 60/159,176 (CIP)  
Filed on 13 October 1999 (13.10.1999)  
(71) Applicant (for all designated States except US): SE-QUENOM, INC. [US/US]; 3595 John Hopkins Court, San Diego, CA 92121 (US).  
(72) Inventors; and  
(75) Inventors/Applicants (for US only): BRAUN, Andreas [DE/US]; 11237-6 Carmel Creek Road, San Diego, CA 92130 (US). KÖSTER, Hubert [DE/CH]; Via Delle Scuole 1, CH-6900 Lugano-Cassarate (CH). VAN DEN BOOM, Dirk [DE/DE]; Eppendorfer Weg 205 D, D-20253 Hamburg (DE). PING, Yip [US/US]; 3641 Copley Avenue, San Diego, CA 92116 (US). RODI,

[Continued on next page]

(54) Title: METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING POLYMORPHIC GENETIC MARKERS



(57) Abstract: Process and methods for creating a database of genomic samples from healthy human donors, methods that use the database to identify and correlate polymorphic genetic markers and other markers with diseases and conditions are provided.

WO 01/027857 A3



Charlie [US/US]: 13823 Recuerdo Drive, Del Mar, CA 92014 (US). **HE, Liyan** [CN/US]: 10948 Creek Bridge Place, San Diego, CA 92128 (US). **CHIU, Norman** [CA/US]: 1128 Caminito Alvarez, San Diego, CA 92126 (US). **JURINKE, Christian** [DE/DE]: Rombergstrasse 22, 20255 Hamburg (DE).

patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— with international search report

(74) **Agents:** SEIDMAN, Stephanie, L. et al.: Heller Ehrman White & McAuliffe, Suite 700, 4250 Executive Square, La Jolla, CA 92037 (US).

(88) **Date of publication of the international search report:**  
3 October 2002

(81) **Designated States (national):** AE, AG, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(48) **Date of publication of this corrected version:**  
5 December 2002

(15) **Information about Correction:**  
see PCT Gazette No. 49/2002 of 5 December 2002, Section II

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**METHODS FOR GENERATING DATABASES AND DATABASES FOR  
IDENTIFYING POLYMORPHIC GENETIC MARKERS**

**RELATED APPLICATIONS**

Benefit of priority to the following applications is claimed herein:

- 5 U.S. provisional application Serial No. 60/217,658 to Andreas Braun, Hubert Koster; Dirk Van den Boom, filed July 10, entitled "METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING POLYMORPHIC GENETIC MARKERS"; U.S. provisional application Serial No. 60/159,176 to Andreas Braun, Hubert Koster, Dirk Van den Boom, filed October 10 13, 1999, entitled "METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING POLYMORPHIC GENETIC MARKERS"; U.S. provisional application Serial No. 60/217,251, filed July 10, 2000, to Andreas Braun, entitled "POLYMORPHIC KINASE ANCHOR PROTEIN GENE SEQUENCES, POLYMORPHIC KINASE ANCHOR PROTEINS AND METHODS OF DETECTING 15 POLYMORPHIC KINASE ANCHOR PROTEINS AND NUCLEIC ACIDS ENCODING THE SAME"; and U.S. application Serial No. 09/663,968, to Ping Yip, filed September 19, 2000, entitled "METHOD AND DEVICE FOR IDENTIFYING A BIOLOGICAL SAMPLE."

- Where permitted that above-noted applications and provisional 20 applications are incorporated by reference in their entirety.

**FIELD OF THE INVENTION**

- Process and methods for creating a database of genomic samples from healthy human donors. Methods that use the database to identify and correlate with polymorphic genetic markers and other markers with diseases and 25 conditions are provided.

**BACKGROUND**

- Diseases in all organisms have a genetic component, whether inherited or resulting from the body's response to environmental stresses, such as viruses and toxins. The ultimate goal of ongoing genomic research is to use this 30 information to develop new ways to identify, treat and potentially cure these diseases. The first step has been to screen disease tissue and identify genomic changes at the level of individual samples. The identification of these "disease"

markers has then fueled the development and commercialization of diagnostic tests that detect these errant genes or polymorphisms. With the increasing numbers of genetic markers, including single nucleotide polymorphisms (SNPs), microsatellites, tandem repeats, newly mapped introns and exons, the challenge  
5 to the medical and pharmaceutical communities is to identify genotypes which not only identify the disease but also follow the progression of the disease and are predictive of an organism's response to treatment.

Currently the pharmaceutical and biotechnology industries find a disease and then attempt to determine the genomic basis for the disease. This approach  
10 is time consuming and expensive and in many cases involves the investigator guessing as to what pathways might be involved in the disease.

#### Genomics

Presently the two main strategies employed in analyzing the available genomic information are the technology driven reverse genetics brute force  
15 strategy and the knowledge-based pathway oriented forward genetics strategy. The brute force approach yields large databases of sequence information but little information about the medical or other uses of the sequence information. Hence this strategy yields intangible products of questionable value. The knowledge-based strategy yields small databases that contain a lot of  
20 information about medical uses of particular DNA sequences and other products in the pathway and yield tangible products with a high value.

#### Polymorphisms

Polymorphisms have been known since 1901 with the identification of blood types. In the 1950's they were identified on the level of proteins using  
25 large population genetic studies. In the 1980's and 1990's many of the known protein polymorphisms were correlated with genetic loci on genomic DNA. For example, the gene dose of the apolipoprotein E type 4 allele was correlated with the risk of Alzheimer's disease in late onset families (see, *e.g.*, Corder *et al.* (1993) *Science* 261: 921-923; mutation in blood coagulation factor V was  
30 associated with resistance to activated protein C (see, *e.g.*, Bertina *et al.* (1994) *Nature* 369:64-67); resistance to HIV-1 infection has been shown in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene (see,

e.g., Samson *et al.* (1996) *Nature* 382:722-725); and a hypermutable tract in antigen presenting cells (APC, such as macrophages), has been identified in familial colorectal cancer in individuals of Ashkenzi jewish background (see, e.g., Laken *et al.* (1997) *Nature Genet.* 17:79-83). There may be more than three million polymorphic sites in the human genome. Many have been identified, but not yet characterized or mapped or associated with a marker.

#### Single nucleotide polymorphisms (SNPs)

Much of the focus of genomics has been in the identification of SNPs, which are important for a variety of reasons. They allow indirect testing (association of haplotypes) and direct testing (functional variants). They are the most abundant and stable genetic markers. Common diseases are best explained by common genetic alterations, and the natural variation in the human population aids in understanding disease, therapy and environmental interactions.

Currently, the only available method to identify SNPs in DNA is by sequencing, which is expensive, difficult and laborious. Furthermore, once a SNP is discovered it must be validated to determine if it is a real polymorphism and not a sequencing error. Also, discovered SNPs must then be evaluated to determine if they are associated with a particular phenotype. Thus, there is a need to develop new paradigms for identifying the genomic basis for disease and markers thereof. Therefore, it is an object herein to provide methods for identifying the genomic basis of disease and markers thereof.

#### SUMMARY

Databases and methods using the databases are provided herein. The databases comprise sets of parameters associated with subjects in populations selected only on the basis of being healthy (*i.e.*, where the subjects are mammals, such as humans, they are selected based upon apparent health and no detectable infections). The databases can be sorted based upon one or more of the selected parameters.

The databases are preferably relational databases, in which an index that represents each subject serves to relate parameters, which are the data, such as age, ethnicity, sex, medical history, etc. and ultimately genotypic information,

that was inputted into and stored in the database. The database can then be sorted according to these parameters. Initially, the parameter information is obtained from a questionnaire answered by each subject from whom a body tissue or body fluid sample is obtained. As additional information about each sample is obtained, this information can be entered into the database and can serve as a sorting parameter.

The databases obtained from healthy individuals have numerous uses, such as correlating known polymorphisms with a phenotype or disease. The databases can be used to identify alleles that are deleterious, that are beneficial, and that are correlated with diseases.

For purposes herein, genotypic information can be obtained by any method known to those of skill in the art, but is preferably obtained using mass spectrometry.

Also provided herein, is a new use for existing databases of subjects and genotypic and other parameters, such as age, ethnicity, race, and gender. Any database can be sorted according to the methods herein and alleles that exhibit statistically significant correlations with any of the sorting parameters can be identified. It is noted, however, is noted, that the databases provided herein and randomly selected databases will perform better in these methods, since disease-based databases suffer numerous limitations, including their relatively small size, the homogeneity of the selected disease population, and the masking effect of the polymorphism associated with the markers for which the database was selected. Hence, the healthy database provided herein, provides advantages not heretofore recognized or exploited. However, the methods provided herein can be used with a selected database, including disease-based databases, with or without sorting for the discovery and correlation of polymorphisms. In addition, the databases provided herein represent a greater genetic diversity than the unselected databases typically utilized for the discovery of polymorphisms and thus allow for the enhanced discovery and correlation of polymorphisms.

The databases provided herein can be used for taking an identified polymorphism, and ascertaining whether it changes in frequency when the data is sorted according to a selected parameter.

One use of these methods is correlating a selected marker with a particular parameter by following the occurrence of known genetic markers and then, having made this correlation, determining or identifying correlations with diseases. Examples of this use are p53 and Lipoprotein Lipase polymorphism.

- 5 As exemplified herein, known markers are shown to have particular correlation with certain groups, such as a particular ethnicity or race or one sex. Such correlations will then permit development of better diagnostic tests and treatment regimens.

- These methods are valuable for identifying one or more genetic markers  
10 whose frequency changes within the population as a function of age, ethnic group, sex or some other criteria. This can allow the identification of previously unknown polymorphisms and ultimately a gene or pathway involved in the onset and progression of disease.

- The databases and methods provided herein permit, among other things,  
15 identification of components, particularly key components, of a disease process by understanding its genetic underpinnings and also permit an understanding of processes, such as individual drug responses. The databases and methods provided herein also can be used in methods involving elucidation of pathological pathways, in developing new diagnostic assays, identifying new potential drug  
20 targets, and in identifying new drug candidates.

The methods and databases can be used with experimental procedures, including, but are not limited to, *in silico* SNP identification, *in vitro* SNP identification/verification, genetic profiling of large populations, and in biostatistical analyses and interpretations.

- 25 Also provided herein, are combinations that contain a database provided herein and a biological sample from a subject in the database, and preferably biological samples from all subjects or a plurality of subjects in the database. Collections of the tissue and body fluid samples are also provided.

- Also, provided herein, are methods for determining a genetic marker that  
30 correlates with age, comprising identifying a polymorphism and determining the frequency of the polymorphism with increasing age in a healthy population.

Further provided herein are methods for determining whether a genetic marker correlates with susceptibility to morbidity, early mortality, or morbidity and early mortality, comprising identifying a polymorphism and determining the frequency of the polymorphism with increasing age in a healthy population.

- 5       Any of the methods herein described can be used out in a multiplex format.

- Also provided are an apparatus and process for accurately identifying genetic information. It is another object of the herein that genetic information be extracted from genetic data in a highly automated manner. Therefore, to  
10       overcome the deficiencies in the known conventional systems, a method and apparatus for identifying a biological sample is proposed.

- Briefly, the method and system for identifying a biological sample generates a data set indicative of the composition of the biological sample. In a particular example, the data set is DNA spectrometry data received from a mass  
15       spectrometer. The data set is denoised, and a baseline is deleted. Since possible compositions of the biological sample may be known, expected peak areas may be determined. Using the expected peak areas, a residual baseline is generated to further correct the data set. Probable peaks are then identifiable in the corrected data set, which are used to identify the composition of the  
20       biological sample. In a disclosed example, statistical methods are employed to determine the probability that a probable peak is an actual peak, not an actual peak, or that the data too inconclusive to call.

- Advantageously, the method and system for identifying a biological sample accurately makes composition calls in a highly automated manner. In  
25       such a manner, complete SNP profile information, for example, may be collected efficiently. More importantly, the collected data is analyzed with highly accurate results. For example, when a particular composition is called, the result may be relied upon with great confidence. Such confidence is provided by the robust computational process employed .

- 30       DESCRIPTION OF THE DRAWINGS

Figure 1 depicts an exemplary sample bank. Panel 1 shows the samples as a function of sex and ethnicity. Panel 2 shows the caucasians as a function of age. Panel 3 shows the Hispanics as a function of age.

Figures 2A and 2C show an age- and sex-distribution of the 291S allele of the lipoprotein lipase gene in which a total of 436 males and 589 females were investigated. Figure 2B shows an age distribution for the 436 males.

Figure 3 is an exemplary questionnaire for population-based sample banking.

Figure 4 depicts processing and tracking of blood sample components.

Figure 5 depicts the allelic frequency of "sick" alleles and "healthy" alleles as a function of age. It is noted that the relative frequency of healthy alleles increases in a population with increasing age.

Figure 6 depicts the age-dependent distribution of ApoE genotypes (see, Schächter *et al.* (1994) *Nature Genetics* 6:29-32).

Figure 7A-D depicts age-related and genotype frequency of the p53 (tumor suppressor) codon 72 among the caucasian population in the database. \*R72 and \*P72 represent the frequency of the allele in the database population. R72, R72P, and P72 represent the genotypes of the individuals in the population. The frequency of the homozygous P72 allele drops from 6.7% to 3.7% with age.

Figure 8 depicts the allele and genotype frequencies of the p21 S31R allele as a function of age.

Figure 9 depicts the frequency of the FVII Allele 353Q in pooled versus individual samples.

Figure 10 depicts the frequency of the CETP (cholesterol ester transfer protein) allele in pooled versus individual samples

Figure 11 depicts the frequency of the plasminogen activator inhibitor-1 (PAI-1) 5G in pooled versus individual samples

Figure 12 shows mass spectra of the samples and the ethnic diversity of the PAI-1 alleles.

Figure 13 shows mass spectra of the samples and the ethnic diversity of the CETP 405 alleles.

Figure 14 shows mass spectra of the samples and the ethnic diversity of the Factor VII 353 alleles.

Figure 15 shows ethnic diversity of PAI-1, CETP and Factor VII using the pooled DNA samples.

5        Figure 16 shows the p53-Rb pathway and the relationships among the various factors in the pathway.

Figure 17, which is a block diagram of a computer constructed to provide and process the databases described herein, depicts a typical computer system for storing and sorting the databases provided herein and practicing the methods  
10        provided herein.

Figure 18 is a flow diagram that illustrates the processing steps performed using the computer illustrated in Figure 17, to maintain and provide access to the databases for identifying polymorphic genetic markers.

Figure 19 is a histogram showing the allele and genotype distribution in  
15        the age and sex stratified Caucasian population for the AKAP10-1 locus. Bright green bars show frequencies in individuals younger than 40 years. Dark green bars show frequencies in individuals older than 60 years.

Figure 20 is a histogram showing the allele and genotype distribution in the age and sex stratified Caucasian population for the AKAP10-5 locus. Bright  
20        green bars show frequencies in individuals younger than 40 years; dark green bars show frequencies in individuals older than 60 years.

Figure 21 is a histogram showing the allele and genotype distribution in the age and sex stratified Caucasian population for the h-msrA locus. Genotype difference between male age groups is significant. Bright green bars show  
25        frequencies in individuals younger than 40 years. Dark green bars show frequencies in individuals older than 60 years.

Figure 22A-D is a sample data collection questionnaire used for the healthy database.

Figure 23 is a flowchart showing processing performed by the computing  
30        device of Figure 24 when performing genotyping of sense strands and antisense strands from assay fragments.

Figure 24 is a block diagram showing a system in accordance with the present invention;

Figure 25 is a flowchart of a method of identifying a biological sample in accordance with the present invention;

5        Figure 26 is a graphical representation of data from a mass spectrometer;  
      Figure 27 is a diagram of wavelet transformation of mass spectrometry data;

      Figure 28 is a graphical representation of wavelet stage 0 hi data;

      Figure 29 is a graphical representation of stage 0 noise profile;

10       Figure 30 is a graphical representation of generating stage noise standard deviations;

      Figure 31 is a graphical representation of applying a threshold to data stages;

      Figure 32 is a graphical representation of a sparse data set;

15       Figure 33 is a formula for signal shifting;

      Figure 34 is a graphical representation of a wavelet transformation of a denoised and shifted signal;

      Figure 35 is a graphical representation of a denoised and shifted signal;

      Figure 36 is a graphical representation of removing peak sections;

20       Figure 37 is a graphical representation of generating a peak free signal ;

      Figure 38 is a block diagram of a method of generating a baseline correction;

      Figure 39 is a graphical representation of a baseline and signal;

      Figure 40 is a graphical representation of a signal with baseline removed;

25       Figure 41 is a table showing compressed data;

      Figure 42 is a flowchart of method for compressing data;

      Figure 43 is a graphical representation of mass shifting;

      Figure 44 is a graphical representation of determining peak width;

      Figure 45 is a graphical representation of removing peaks;

30       Figure 46 is a graphical representation of a signal with peaks removed;

      Figure 47 is a graphical representation of a residual baseline;

Figure 48 is a graphical representation of a signal with residual baseline removed;

Figure 49 is a graphical representation of determining peak height;

Figure 50 is a graphical representation of determining signal-to-noise for  
5 each peak;

Figure 51 is a graphical representation of determining a residual error for each peak;

Figure 52 is a graphical representation of peak probabilities;

Figure 53 is a graphical representation of applying an allelic ratio to peak  
10 probability;

Figure 54 is a graphical representation of determining peak probability

Figure 55 is a graphical representation of calling a genotype;

Figure 56 is a flowchart showing a statistical procedure for calling a  
genotype;

15 Figure 57 is a flowchart showing processing performed by the computing device of Figure 1 when performing standardless genotyping; and

Figure 58 is graphical representation of applying an allelic ratio to peak probability for standardless genotype processing.

#### **DETAILED DESCRIPTION**

##### **20 Definitions**

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art to which this invention belongs. All patents, applications, published applications and other publications and sequences from GenBank and other  
25 databases referred to herein throughout the disclosure are incorporated by reference in their entirety.

As used herein, a biopolymer includes, but is not limited to, nucleic acid, proteins, polysaccharides, lipids and other macromolecules. Nucleic acids include DNA, RNA, and fragments thereof. Nucleic acids may be derived from  
30 genomic DNA, RNA, mitochondrial nucleic acid, chloroplast nucleic acid and other organelles with separate genetic material.

As used herein, morbidity refers to conditions, such as diseases or disorders, that compromise the health and well-being of an organism, such as an animal. Morbidity susceptibility or morbidity-associated genes are genes that, when altered, for example, by a variation in nucleotide sequence, facilitate the expression of a specific disease clinical phenotype. Thus, morbidity susceptibility genes have the potential, upon alteration, of increasing the likelihood or general risk that an organism will develop a specific disease.

As used herein, mortality refers to the statistical likelihood that an organism, particularly an animal, will not survive a full predicted lifespan. Hence, a trait or a marker, such as a polymorphism, associated with increased mortality is observed at a lower frequency in older than younger segments of a population.

As used herein, a polymorphism, e.g. genetic variation, refers to a variation in the sequence of a gene in the genome amongst a population, such as allelic variations and other variations that arise or are observed. Thus, a polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. These differences can occur in coding and non-coding portions of the genome, and can be manifested or detected as differences in nucleic acid sequences, gene expression, including, for example transcription, processing, translation, transport, protein processing, trafficking, DNA synthesis, expressed proteins, other gene products or products of biochemical pathways or in post-translational modifications and any other differences manifested amongst members of a population. A single nucleotide polymorphism (SNP) refers to a polymorphism that arises as the result of a single base change, such as an insertion, deletion or change in a base.

A polymorphic marker or site is the locus at which divergence occurs. Such site may be as small as one base pair (an SNP). Polymorphic markers include, but are not limited to, restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats and other repeating patterns, simple sequence repeats and insertional elements, such as Alu. Polymorphic forms also are manifested as different mendelian

alleles for a gene. Polymorphisms may be observed by differences in proteins, protein modifications, RNA expression modification, DNA and RNA methylation, regulatory factors that alter gene expression and DNA replication, and any other manifestation of alterations in genomic nucleic acid or organelle nucleic acids.

- 5 As used herein, a healthy population, refers to a population of organisms, including but are not limited to, animals, bacteria, viruses, parasites, plants, eubacteria, and others, that are disease free. The concept of disease-free is a function of the selected organism. For example, for mammals it refers to a subject not manifesting any disease state. Practically a healthy subject, when
- 10 human, is defined as human donor who passes blood bank criteria to donate blood for eventual use in the general population. These criteria are as follows: free of detectable viral, bacterial, mycoplasma, and parasitic infections; not anemic; and then further selected based upon a questionnaire regarding history (see Figure 3). Thus, a healthy population represents an unbiased population of
- 15 sufficient health to donate blood according to blood bank criteria, and not further selected for any disease state. Typically such individuals are not taking any medications. For plants, for example, it is a plant population that does not manifest diseases pathology associated with plants. For bacteria it is a bacterial population replicating without environmental stress, such as selective agents,
- 20 heat and other pathogens.

- As used herein, a healthy database (or healthy patient database) refers to a database of profiles of subjects that have not been pre-selected for any particular disease. Hence, the subjects that serve as the source of data for the database are selected, according to predetermined criteria, to be healthy. In
- 25 contrast to other such databases that have been pre-selected for subjects with a particular disease or other characteristic, the subjects for the database provided herein are not so-selected. Also, if the subjects do manifest a disease or other condition, any polymorphism discovered or characterized should be related to an independent disease or condition. In a preferred embodiment, where the
- 30 subjects are human, a healthy subject manifests no disease symptoms and meets criteria, such as those set by blood banks for blood donors.

Thus, the subjects for the database are a population of any organism, including, but are not limited to, animals, plants, bacteria, viruses, parasites and any other organism or entity that has nucleic acid. Among preferred subjects are mammals, preferably, although not necessarily, humans. Such a database can  
5 capture the diversity of the a population, thus providing for discovery of rare polymorphisms.

As used herein, a profile refers to information relating to, but not limited to and not necessarily including all of, age, sex, ethnicity, disease history, family history, phenotypic characteristics, such as height and weight and other relevant  
10 parameters. A sample collect information form is shown in Figure 22, which illustrates profile intent.

As used herein, a disease state is a condition or abnormality or disorder that may be inherited or result from environmental stresses, such as toxins, bacterial, fungal and viral infections.

15 As used herein, set of non-selected subjects means that the subjects have not been pre-selected to share a common disease or other characteristic. They can be selected to be healthy as defined herein.

As used herein, a phenotype refers to a set of parameters that includes any distinguishable trait of an organism. A phenotype can be physical traits and  
20 can be, in instances in which the subject is an animal, a mental trait, such as emotional traits. Some phenotypes can be determined by observation elicited by questionnaires (see, *e.g.*, Figures 3 and 22) or by referring to prior medical and other records. For purposes herein, a phenotype is a parameter around which the database can be sorted.

25 As used herein, a parameter is any input data that will serve as a basis for sorting the database. These parameters will include phenotypic traits, medical histories, family histories and any other such information elicited from a subject or observed about the subject. A parameter may describe the subject, some historical or current environmental or social influence experienced by the  
30 subject, or a condition or environmental influence on someone related to the subject. Paramaters include, but are not limited to, any of those described herein, and known to those of skill in the art.

As used herein, haplotype refers to two or more polymorphisms located on a single DNA strand. Hence, haplotyping refers to identification of two or more polymorphisms on a single DNA strand. Haplotypes can be indicative of a phenotype. For some disorders a single polymorphism may suffice to indicate a trait; for others a plurality (i.e., a haplotype) may be needed. Haplotyping can be performed by isolating nucleic acid and separating the strands. In addition, when using enzymes such as certain nucleases, that produce, different size fragments from each strand, strand separation is not needed for haplotyping.

As used herein, used herein, pattern with reference to a mass spectrum or mass spectrometric analyses, refers to a characteristic distribution and number of signals (such peaks or digital representations thereof).

As used herein, signal in the context of a mass spectrum and analysis thereof refers to the output data, which the number or relative number of molecules having a particular mass. Signals include "peaks" and digital representations thereof.

As used herein, adaptor, when used with reference to haplotyping using Fen ligase, refers to a nucleic acid that specifically hybridizes to a polymorphism of interest. An adaptor can be partially double-stranded. An adaptor complex is formed when an adaptor hybridizes to its target.

As used herein, a target nucleic acid refers to any nucleic acid of interest in a sample. It can contain one or more nucleotides.

As used herein, standardless analysis refers to a determination based upon an internal standard. For example, the frequency of a polymorphism can be determined herein by comparing signals within a single mass spectrum.

As used herein, amplifying refers to means for increasing the amount of a bipolymer, especially nucleic acids. Based on the 5' and 3' primers that are chosen, amplification also serves to restrict and define the region of the genome which is subject to analysis. Amplification can be done by any means known to those skilled in the art, including use of the polymerase chain reaction (PCR) etc.

Amplification, e.g., PCR must be done quantitatively when the frequency of polymorphism is required to be determined.

As used herein, cleaving refers to non-specific and specific fragmentation of a biopolymer.

As used herein, multiplexing refers to the simultaneous detection of more than one polymorphism. Methods for performing multiplexed reactions,

- 5 particularly in conjunction with mass spectrometry are known (see, *e.g.*, U.S. Patent Nos. 6,043,031, 5,547,835 and International PCT application No. WO 97/37041).

- As used herein, reference to mass spectrometry encompasses any suitable mass spectrometric format known to those of skill in the art. Such formats
- 10 include, but are not limited to, Matrix-Assisted Laser Desorption/Ionization, Time-of-Flight (MALDI-TOF), Electrospray (ES), IR-MALDI (see, *e.g.*, published International PCT application No.99/57318 and U.S. Patent No. 5,118,937), Ion Cyclotron Resonance (ICR), Fourier Transform and combinations thereof. MALDI, particular UV and IR, are among the preferred formats.

- 15 As used herein, mass spectrum refers to the presentation of data obtained from analyzing a biopolymer or fragment thereof by mass spectrometry either graphically or encoded numerically.

- As used herein, a blood component is a component that is separated from blood and includes, but is not limited to red blood cells and platelets, blood
- 20 clotting factors, plasma, enzymes, plasminogen, immunoglobulins. A cellular blood component is a component of blood, such as a red blood cell, that is a cell. A blood protein is a protein that is normally found in blood. Examples of such proteins are blood factors VII and VIII. Such proteins and components are well-known to those of skill in the art.

- 25 As used herein, plasma can be prepared by any method known to those of skill in the art. For example, it can be prepared by centrifuging blood at a force that pellets the red cells and forms an interface between the red cells and the buffy coat, which contains leukocytes, above which is the plasma. For example, typical platelet concentrates contain at least about 10% plasma.

- 30 Blood may be separated into its components, including, but not limited to, plasma, platelets and red blood cells by any method known to those of skill in the art. For example, blood can be centrifuged for a sufficient time and at a

sufficient acceleration to form a pellet containing the red blood cells. Leukocytes collect primarily at the interface of the pellet and supernatant in the buffy coat region. The supernatant, which contains plasma, platelets, and other blood components, may then be removed and centrifuged at a higher acceleration,  
5 whereby the platelets pellet.

As used herein, p53 is a cell cycle control protein that assesses DNA damage and acts as a transcription factor regulation gene which control cell growth, DNA repair and apoptosis. The p53 mutations have been found in a wide variety of different cancers, including all of the different types of leukemia,  
10 with varying frequency. The loss of normal p53 functions results in genomic instability and uncontrolled growth of the host cell.

As used herein, p21 is a cyclin-dependent kinase inhibitor, associated with G1 phase arrest of normal cells. Expression triggers apoptosis or programmed cell death and has been associated with Wilms' tumor, a pediatric  
15 kidney cancer.

As used herein, Factor VII is a serine protease involved the extrinsic blood coagulation cascade. This factor is activated by thrombin and works with tissue factor (Factor III) in the processing of Factor X to Factor Xa. Evidence has supported an association between polymorphisms in the gene and increase  
20 Factor VII activity which can result in an elevated risk of ischemic cardiovascular disease including myocardial infarction.

As used herein, a relational database stores information in a form representative of matrices, such as two-dimensional tables, including rows and columns of data, or higher dimensional matrices. For example, in one  
25 embodiment, the relational database has separate tables each with a parameter. The tables are linked with a record number, which also acts as an index. The database can be searched or sorted by using data in the tables and is stored in any suitable storage medium, such as floppy disk, CD rom disk, hard drive or other suitable medium.

30 As used herein, a bar codes refers any array of optically readable marks of any desired size and shape that are arranged in a reference context or frame of, preferably, although not necessarily, one or more columns and one or more

rows. For purposes herein, the bar code refers to any symbology, not necessary "bar" but may include dots, characters or any symbol or symbols.

As used herein, symbology refers to an identifier code or symbol, such as a bar code, that is linked to a sample. The index will reference each such  
5 symbology. The symbology is any code known or designed by the user. The symbols are associated with information stored in the database. For example, each sample can be uniquely identified with an encoded symbology. The parameters, such as the answers to the questions and subsequent genotypic and other information obtained upon analysis of the samples is included in the  
10 database and associated with the symbology. The database is stored on any suitable recording medium, such as a hard drive, a floppy disk, a tape, a CD ROM, a DVD disk and any other suitable medium.

#### DATABASES

Human genotyping is currently dependent on collaborations with  
15 hospitals, tissues banks and research institutions that provide samples of disease tissue. This approach is based on the concept that the onset and/or progression of diseases can be correlated with the presence of a polymorphisms or other genetic markers. This approach does not consider that disease correlated with the presence of specific markers and the absence of specific markers. It is  
20 shown herein that identification and scoring of the appearance and disappearance of markers is possible only if these markers are measured in the background of healthy subjects where the onset of disease does not mask the change in polymorphism occurrence. Databases of information from disease populations suffer from small sample size, selection bias and heterogeneity. The  
25 databases provided herein from healthy populations solve these problems by permitting large sample bands, simple selection methods and diluted heterogeneity.

Provided herein are first databases of parameters, associated with non-selected, particularly healthy, subjects. Also provided are combinations of the  
30 databases with indexed samples obtained from each of the subjects. Further provided are databases produced from the first databases. These contain in addition to the original parameters information, such as genotypic information,

including, but are not limited to, genomic sequence information, derived from the samples.

The databases, which are herein designated healthy databases, are so-designated because they are not obtained from subjects pre-selected for a particular disease. Hence, although individual members may have a disease, the collection of individuals is not selected to have a particular disease.

The subjects from whom the parameters are obtained comprise either a set of subjects who are randomly selected across, preferably, all populations, or are pre-selected to be disease-free or healthy. As a result, the database is not selected to be representative of any pre-selected phenotype, genotype, disease or other characteristic. Typically the number of subjects from which the database is prepared is selected to produce statistically significant results when used in the methods provided herein. Preferably, the number of subjects will be greater than 100, more preferably greater than 200, yet more preferably greater than 1000. The precise number can be empirically determined based upon the frequency of the parameter(s) that be used to sort the database. Generally the population can have at least 50, at least 100, at least 200, at least 500, at least 1000, at least 5000 or at least 10,000 or more subjects.

Upon identification of a collection of subjects, information about each subject is recorded and associated with each subject as a database. The information associated with each of the subjects, includes, but is not limited to, information related to historical characteristics of the subjects, phenotypic characteristics and also genotypic characteristics, medical characteristics and any other traits and characteristics about the subject that can be determined. This information will serve as the basis for sorting the database.

In an exemplary embodiment, the subjects are mammals, such as humans, and the information relates to one or more of parameters, such as age, sex, medical history, ethnicity and any other factor. Such information, when the animals are humans, for example, can be obtained by a questionnaire, and by observations about the individual, such as hair color, eye color and other characteristics. Genotypic information will be obtained from tissue or other body and body fluid samples from the subject.

The healthy genomic database can include profiles and polymorphisms from healthy individuals from a library of blood samples where each sample in the library is an individual and separate blood or other tissue sample. Each sample in the database is profiled as to the sex, age, ethnic group, and disease history of the donor.

The databases are generated by first identifying healthy populations of subjects and obtaining information about each subject that will serve as the sorting parameters for the database. This information is preferably entered into a storage medium, such as the memory of a computer.

The information obtained about each subject in a population used for generating the database is stored in a computer memory or other suitable storage medium. The information is linked to an identifier associated with each subject. Hence the database will identify a subject, for example by a datapoint representative of a bar code, and then all information, such as the information from a questionnaire, regarding the individual is associated with the datapoint. As the information is collected the database is generated.

Thus, for example, profile information, such as subject histories obtained from questionnaires, is collected in the database. The resulting database can be sorted as desired, using standard software, such as by age, sex and/or ethnicity. An exemplary questionnaire for subjects from whom samples are to be obtained is shown in Figures 22A-D. Each questionnaire preferably is identified by a bar code, particularly a machine readable bar code for entry into the database. After a subject provides data and is deemed to be healthy (*i.e.*, meets standards for blood donation), the data in the questionnaire is entered into the database and is associated with the bar code. A tissue, cell or blood sample is obtained from the subject.

Figure 4 exemplifies processing and tracking of blood sample components. Each component is tracked with a bar code, dated, is entered into the database and associated with the subject and the profile of the subject. Typically, the whole blood is centrifuged to produce plasma, red blood cells (which pellet) and leukocytes found in the buffy coat which layers in between.

Various samples are obtained and coded with a bar code and stored for use as needed.

Samples are collected from the subjects. The samples include, but are not limited to, tissues, cells, and fluids, such as nucleic acid, blood, plasma, 5 amniotic fluid, synovial fluid, urine, saliva, aqueous humor, sweat, sperm samples and cerebral spinal fluid. It is understood that the particular set of samples depends upon the organisms in the population.

Once samples are obtained the collection can be stored and, in preferred embodiments, each sample is indexed with an identifier, particularly a machine 10 readable code, such as a bar code. For analyses, the samples or components of the samples, particularly biopolymers and small molecules, such as nucleic acids and/or proteins and metabolites, are isolated.

After samples are analyzed, this information is entered into the database in the memory of the storage medium and associated with each subject. This 15 information includes, but is not limited to, genotypic information. Particularly, nucleic acid sequence information and other information indicative of polymorphisms, such as masses of PCR fragments, peptide fragment sequences or masses, spectra of biopolymers and small molecules and other indicia of the structure or function of a gene, gene product or other marker from which the 20 existence of a polymorphism within the population can be inferred.

In an exemplary embodiment, a database can be derived from a collection of blood samples. For example, Figure 1 (see, also Figure 10) shows the status of a collection of over 5000 individual samples. The samples were processed in the laboratory following SOP (standard operating procedure) guidelines. Any 25 standard blood processing protocol may be used.

For the exemplary database described herein, the following criteria were used to select subjects:

No testing is done for infectious agents.

Age: At least 17 years old

30 Weight: Minimum of 110 pounds

Permanently Disqualified:

History of hepatitis (after age 11)

Leukemia Lymphoma

Human immunodeficiency virus (HIV), AIDS

Chronic kidney disease

**Temporarily Disqualified:**

- 5       Pregnancy - until six weeks after delivery, miscarriage or abortion  
Major surgery or transfusions - for one year  
Mononucleosis - until complete recovery  
Prior whole blood donation - for eight weeks  
Antibiotics by injection for one week; by mouth, for forty-eight hours,  
10       except antibiotics for skin complexion;

**5 year Deferment:**

Internal cancer and skin cancer if it has been removed, is healed and  
there is no recurrence

- These correspond to blood bank criteria for donating blood and represent a  
15       healthy population as defined herein for a human healthy database.

**Structure of the database**

- Any suitable database structure and format known to those of skill in the  
art may be employed. For example, a relational database is a preferred format in  
which data is stored as matrices or tables of the parameters linked by an indexer  
20       that identifies each subject. Software for preparing and manipulating, including  
sorting the database, can be readily developed or adapted from commercially  
available software, such as Microsoft Access.

**Quality control**

- Quality control procedures can be implemented. For example, after  
25       collection of samples, the quality of the collection in the bank can be assessed.  
For example, mix-up of samples can be checked by testing for known markers,  
such as sex. After samples are separated by ethnicity, samples are randomly  
tested for a marker associated with a particular ethnicity, such as HLA DQA1  
group specific component, to assess whether the samples have been properly  
30       sorted by ethnic group. An exemplary sample bank is depicted in Figure 4.

### Obtaining genotypic data and other parameters for the database

After informational and historical parameters are entered into the database, material from samples obtained from each subject, is analyzed. Analyzed material include proteins, metabolites, nucleic acids, lipids and any  
5 other desired constituent of the material. For example, nucleic acids, such as genomic DNA, can be analyzed by sequencing.

Sequencing can be performed using any method known to those of skill in the art. For example, if a polymorphism is identified or known, and it is desired to assess its frequency or presence among the subjects in the database, the  
10 region of interest from each sample can be isolated, such as by PCR or restriction fragments, hybridization or other suitable method known to those of skill in the art and sequenced. For purposes herein, sequencing analysis is preferably effected using mass spectrometry (see, *e.g.*, U.S. Patent Nos. 5,547,835, 5,622,824, 5,851,765, and 5,928,906). Nucleic acids can also be  
15 sequence by hybridization (see, *e.g.*, U.S. Patent Nos. 5,503,980, 5,631,134, 5,795,714) and including analysis by mass spectrometry (see, U.S. application Serial Nos. 08/419,994 and 09/395,409).

In other detection methods, it is necessary to first amplify prior to identifying the allelic variant. Amplification can be performed, *e.g.*, by PCR  
20 and/or LCR, according to methods known in the art. In one embodiment, genomic DNA of a cell is exposed to two PCR primers and amplification for a number of cycles sufficient to produce the required amount of amplified DNA. In preferred embodiments, the primers are located between 150 and 350 base pairs apart.

Alternative amplification methods include: self sustained sequence  
25 replication (Guatelli, J. C. et al., 1990, Proc. Natl. Acad. Sci. U.S.A. 87:1874-1878), transcriptional amplification system (Kwoh, D. Y. et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:1173-1177), Q-Beta Replicase (Lizardi, P. M. et al., 1988, Bio/Technology 6:1197), or any other nucleic acid amplification  
30 method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially

useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

- Nucleic acids can also be analyzed by detection methods and protocols, particularly those that rely on mass spectrometry (see, *e.g.*, U.S. Patent No. 5,605,798, 6,043,031, allowed copending U.S. application Serial No. 08/744,481, U.S. application Serial No. 08/990,851 and International PCT application No. WO 99/31273, International PCT application No. WO 98/20019). These methods can be automated (see, *e.g.*, copending U.S. application Serial No. 09/285,481 and published International PCT application No. PCT/US00/08111, which describes an automated process line). Preferred among the methods of analysis herein are those involving the primer oligo base extension (PROBE) reaction with mass spectrometry for detection (described herein and elsewhere, see *e.g.*, U.S. Patent No. 6,043,031; see, also U.S. application Serial Nos. 09/287,681, 09/287,682, 09/287,141 and 09/287,679, allowed copending U.S. application Serial No. 08/744,481, International PCT application No. PCT/US97/20444, published as International PCT application No. WO 98/20019, and based upon U.S. application Serial Nos. 08/744,481, 08/744,590, 08/746,036, 08/746,055, 08/786,988, 08/787,639, 08/933,792, 08/746,055, 08/786,988 and 08/787,639; see, also U.S. application Serial No. 09/074,936, U.S. Patent No. 6,024,925, and U.S. application Serial Nos. 08/746,055 and 08/786,988, and published International PCT application No. WO 98/20020)

- A preferred format for performing the analyses is a chip based format in which the biopolymer is linked to a solid support, such as a silicon or silicon-coated substrate, preferably in the form of an array. More preferably, when analyses are performed using mass spectrometry, particularly MALDI, small nanoliter volumes of sample are loaded on, such that the resulting spot is about, or smaller than, the size of the laser spot. It has been found that when this is achieved, the results from the mass spectrometric analysis are quantitative. The area under the signals in the resulting mass spectra are proportional to concentration (when normalized and corrected for background). Methods for preparing and using such chips are described in U.S. Patent No. 6,024,925, co-

pending U.S. application Serial Nos. 08/786,988, 09/364,774, 09/371,150 and 09/297,575; see, also U.S. application Serial No. PCT/US97/20195, which published as WO 98/20020. Chips and kits for performing these analyses are commercially available from SEQUENOM under the trademark MassARRAY.

- 5 MassArray relies on the fidelity of the enzymatic primer extension reactions combined with the miniaturized array and MALDI-TOF (Matrix-Assisted Laser Desorption Ionization-Time of Flight) mass spectrometry to deliver results rapidly. It accurately distinguishes single base changes in the size of DNA fragments associated with genetic variants without tags.

- 10 The methods provided herein permit quantitative determination of alleles. The areas under the signals in the mass spectra can be used for quantitative determinations. The frequency is determined from the ratio of the signal to the total area of all of the spectrum and corrected for background. This is possible because of the PROBE technology as described in the above applications

- 15 incorporated by reference herein.

- Additional methods of analyzing nucleic acids include amplification- based methods including polymerase chain reaction (PCR), ligase chain reaction (LCR), mini-PCR, rolling circle amplification, autocatalytic methods, such as those using Q $\beta$  replicase, TAS, 3SR, and any other suitable method known to those of skill
- 20 in the art.

Other methods for analysis and identification and detection of polymorphisms, include but are not limited to, allele specific probes, Southern analyses, and other such analyses.

- The methods described below provide ways to fragment given amplified
- 25 or non-amplified nucleotide sequences thereby producing a set of mass signals when mass spectrometry is used to analyze the fragment mixtures.

- Amplified fragments are yielded by standard polymerase chain methods (US 4,683,195 and 4,683,202). The fragmentation method involves the use of enzymes that cleave single or double strands of DNA and enzymes that ligate
- 30 DNA. The cleavage enzymes can be glycosylases, nickases, and site-specific and non site-specific nucleases with the most preferred enzymes being glycosylases, nickases, and site-specific nucleases.

### Glycosylase Fragmentation Method

DNA glycosylases specifically remove a certain type of nucleobase from a given DNA fragment. These enzymes can thereby produce abasic sites, which can be recognized either by another cleavage enzyme, cleaving the exposed  
5 phosphate backbone specifically at the abasic site and producing a set of nucleobase specific fragments indicative of the sequence, or by chemical means, such as alkaline solutions and or heat. The use of one combination of a DNA glycosylase and its targeted nucleotide would be sufficient to generate a base specific signature pattern of any given target region.

10 Numerous DNA glycosylases are known. For example, a DNA glycosylase can be uracil-DNA glycosylase (UDG), 3-methyladenine DNA glycosylase, 3-methyladenine DNA glycosylase II, pyrimidine hydrate-DNA glycosylase, FaPy-DNA glycosylase, thymine mismatch-DNA glycosylase, hypoxanthine-DNA glycosylase, 5-Hydroxymethyluracil DNA glycosylase (HmUDG), 5-  
15 Hydroxymethylcytosine DNA glycosylase, or 1,N6-ethenoadenine DNA glycosylase (see, *e.g.*, U.S. Patent Nos. 5,536,649, 5,888, 795, 5,952,176 and 6,099,553, International PCT application Nos. WO 97/03210, WO 99/54501; see, also, Eftedal et al. (1993) Nucleic Acids Res 21:2095-2101, Bjelland and Seeberg (1987) Nucleic Acids Res. 15:2787-2801, Saparbaev et al.  
20 (1995) Nucleic Acids Res. 23:3750-3755, Bessho (1999) Nucleic Acids Res. 27:979-983) corresponding to the enzyme's modified nucleotide or nucleotide analog target. A preferred glycosylase is uracil-DNA glycosylase (UDG).

Uracil, for example, can be incorporated into an amplified DNA molecule by amplifying the DNA in the presence of normal DNA precursor nucleotides  
25 (e.g. dCTP, dATP, and dGTP) and dUTP. When the amplified product is treated with UDG, uracil residues are cleaved. Subsequent chemical treatment of the products from the UDG reaction results in the cleavage of the phosphate backbone and the generation of nucleobase specific fragments. Moreover, the separation of the complementary strands of the amplified product prior to  
30 glycosylase treatment allows complementary patterns of fragmentation to be generated. Thus, the use of dUTP and Uracil DNA glycosylase allows the generation of T specific fragments for the complementary strands, thus providing

information on the T as well as the A positions within a given sequence. Similar to this, a C-specific reaction on both (complementary) strands (i.e. with a C-specific glycosylase) yields information on C as well as G positions within a given sequence if the fragmentation patterns of both amplification strands are  
5 analyzed separately. Thus, with the glycosylase method and mass spectrometry, a full series of A, C, G and T specific fragmentation patterns can be analyzed.

#### Nickase Fragmentation Method

A DNA nickase, or DNase, can be used recognize and cleave one strand  
10 of a DNA duplex. Numerous nickases are known. Among these, for example, are nickase NY2A nickase and NYS1 nickase (Megabase) with the following cleavage sites:

NY2A: 5'...R AG...3'

3'...Y TC...5' where R = A or G and Y = C or T

15 NYS1: 5'... CC[A/G/T]...3'

3'... GG[T/C/A]...5'.

#### Fen-Ligase Fragmentation Method

The Fen-ligase method involves two enzymes: Fen-1 enzyme and a ligase. The Fen-1 enzyme is a site-specific nuclease known as a "flap" endonuclease  
20 (US 5,843,669, 5,874,283, and 6,090,606). This enzymes recognizes and cleaves DNA "flaps" created by the overlap of two oligonucleotides hybridized to a target DNA strand. This cleavage is highly specific and can recognize single base pair mutations, permitting detection of a single homologue from an individual heterozygous at one SNP of interest and then genotyping that  
25 homologue at other SNPs occurring within the fragment. Fen-1 enzymes can be Fen-1 like nucleases e.g. human, murine, and Xenopus XPG enzymes and yeast RAD2 nucleases or Fen-1 endonucleases from, for example, *M. jannaschii*, *P. furiosus*, and *P. woesei*. Among preferred enzymes are the Fen-1 enzymes.

The ligase enzyme forms a phosphodiester bond between two double  
30 stranded nucleic acid fragments. The ligase can be DNA Ligase I or DNA Ligase III (see, e.g., U.S. Patent Nos. US 5,506,137, 5,700,672, 5,858,705 and 5,976,806; see, also, Waga, *et al.* (1994) J. Biol. Chem. 269:10923-10934, Li

et al. (1994) *Nucleic Acids Res.* 22:632-638, Arrand et al. (1986) *J. Biol. Chem.* 261:9079-9082, Lehman (1974) *Science* 186:790-797, Higgins and Cozzarelli (1979) *Methods Enzymol.* 68:50-71, Lasko et al. (1990) *Mutation Res.* 236:277-287, and Lindahl and Barnes (1992) *Ann. Rev. Biochem.* 61:251-281).

- 5           Thermostable ligase (Epicenter Technologies), where "thermostable" denotes that the ligase retains activity even after exposure to temperatures necessary to separate two strands of DNA, are among preferred ligases for use herein.

#### Type IIS Enzyme Fragmentation Method

- 10           Restriction enzymes bind specifically to and cleave double-stranded DNA at specific sites within or adjacent to a particular recognition sequence. These enzymes have been classified into three groups (e.g. Types I, II, and III) as known to those of skill in the art. Because of the properties of type I and type III enzymes, they have not been widely used in molecular biological applications.
- 15           Thus, for this invention type II enzymes are preferred. Of the thousands of restriction enzymes known in the arts, there are 179 different type II specificities. Of the 179 unique type II restriction endonucleases, 31 have a 4-base recognition sequence, 11 have a 5-base recognition sequence, 127 have a 6-base recognition sequence, and 10 have recognition sequences of greater than
- 20           six bases (US 5,604,098). Of category type II enzymes, type IIS is preferred.

          Type IIS enzymes can be *Alw* XI, *Bbv* I, *Bce* 83, *Bpm* I, *Bsg* I, *Bsm* AI, *Bsm* FI, *Bsa* I, *Bcc* I, *Bcg* I, *Ear* I, *Eco* 57I, *Esp* 3I, *Fau* I, *Fok* I, *Gsu* I, *Hga* I, *Mme* I, *Mbo* II, *Sap* I, and the like. The preferred type IIS enzyme is *Fok* I.

- The *Fok* I enzyme endonuclease is an exemplary well characterized
- 25           member of the Type IIS class (see, e.g., U.S. Patent Nos. 5,714,330, 5,604,098, 5,436,150, 6,054,276 and 5,871,911; see, also, Szybalski et al. (1991) *Gene* 100:13-26, Wilson and Murray (1991) *Ann. Rev. Genet.* 25:585-627, Sugisaki et al. (1981) *Gene* 16:73-78, Podhajska and Szalski (1985) *Gene* 40:175-182. *Fok* I recognizes the sequence 5'GGATG-3' and cleaves DNA
- 30           accordingly. Type IIS restriction sites can be introduced into DNA targets by incorporating the site into primers used to amplify such targets. Fragments produced by digestion with *Fok* I are site specific and can be analyzed by mass

spectrometry methods such as MALDI-TOF mass spectrometry, ESI-TOF mass spectrometry, and any other type of mass spectrometry well known to those of skill in the art.

Once a polymorphism has been found to correlate with a parameter  
5 such as age. The possibility of false results due to allelic dropout is examined by doing comparative PCR in an adjacent region of the genome.

#### **Analyses**

In using the database, allelic frequencies can be determined across the population by analyzing each sample in the population individually, determining  
10 the presence or absence of allele or marker of interest in each individual sample, and then determining the frequency of the marker in the population. The database can then be sorted (stratified) to identify any correlations between the allele and a selected parameter using standard statistical analysis. If a correlation is observed, such as a decrease in a particular marker with age or  
15 correlation with sex or other parameter, then the marker is a candidate for further study, such as genetic mapping to identify a gene or pathway in which it is involved. The marker may then be correlated, for example, with a disease. Haplotyping can also be carried out. Genetic mapping can be effected using standard methods and may also require use of databases of others, such as  
20 databases previously determined to be associated with a disorder.

Exemplary analyses have been performed and these are shown in the figures, and discussed herein.

#### **Sample pooling**

It has been found that using the databases provided herein, or any other  
25 database of such information, substantially the same frequencies that were obtained by examining each sample separately can be obtained by pooling samples, such as in batches of 10, 20, 50, 100, 200, 500, 1000 or any other number. A precise number may be determined empirically if necessary, and can be as low as 3.

In one embodiment, the frequency of genotypic and other markers can be obtained by pooling samples. To do this a target population and a genetic variation to be assessed is selected, a plurality of samples of biopolymers are obtained from members of the population, and the biopolymer from which the marker or genotype can be inferred is determined or detected. A comparison of samples tested in pools and individually and the sorted results therefrom are shown in Figure 9, which shows frequency of the factor VII Allele 353Q. Figure 10 depicts the frequency of the CETP Allele CETP in pooled versus individual samples. Figure 15 shows ethnic diversity among various ethnic groups in the database using pooled DNA samples to obtain the data. Figures 12-14 show mass spectra for these samples.

Pooling of test samples has application not only to the healthy databases provided herein, but also to use in gathering data for entry into any database of subjects and genotypic information, including typical databases derived from diseased populations. What is demonstrated herein, is the finding that the results achieved are statistically the same as the results that would be achieved if each sample is analyzed separately. Analysis of pooled samples by a method, such as the mass spectrometric methods provided herein, permits resolution of such data and quantitation of the results.

For factor VII the R53Q acid polymorphism was assessed. In Figure 9, the "individual" data represent allelic frequency observed in 92 individuals reactions. The pooled data represent the allelic frequency of the same 92 individuals pooled into a single probe reaction. The concentration of DNA in the samples of individual donors is 250 nanograms. The total concentration of DNA in the pooled samples is also 250 nanograms, where the concentration of any individual DNA is 2.7 nanograms.

It also was shown that it is possible to reduce the DNA concentration of individuals in a pooled samples from 2.7 nanograms to 0.27 nanograms without any change in the quality of the spectrum or the ability to quantitate the amount of sample detected. Hence low concentrations of sample may be used in the pooling methods.

### Use of the databases and markers identified thereby

The successful use of genomics requires a scientific hypothesis (*i.e.*, common genetic variation, such as a SNP), a study design (*i.e.*, complex disorders), samples and technology, such as the chip-based mass spectrometric analyses (see, *e.g.*, U.S. Patent No. 5,605,798, U.S. Patent No. 5,777,324, 5 U.S. Patent No. 6,043,031, allowed copending U.S. application Serial No. 08/744,481, U.S. application Serial No. 08/990,851, International PCT application No. WO 98/20019, copending U.S. application Serial No. 09/285,481, which describes an automated process line for analyses; see, also, 10 U.S. application Serial Nos. 08/617,256, 09/287,681, 09/287,682, 09/287,141 and 09/287,679, allowed copending U.S. application Serial No. 08/744,481, International PCT application No. PCT/US97/20444, published as International PCT application No. WO 98/20019, and based upon U.S. application Serial Nos. 08/744,481, 08/744,590, 08/746,036, 08/746,055, 08/786,988, 08/787,639, 15 08/933,792, 08/746,055, 09/266,409, 08/786,988 and 08/787,639; see, also U.S. application Serial No. 09/074,936). All of these aspects can be used in conjunction with the databases provided herein and samples in the collection.

The databases and markers identified thereby can be used, for example, for identification of previously unidentified or unknown genetic markers and to 20 identify new uses for known markers. As markers are identified, these may be entered into the database to use as sorting parameters from which additional correlations may be determined.

### Previously unidentified or unknown genetic markers

The samples in the healthy databases can be used to identify new 25 polymorphisms and genetic markers, using any mapping, sequencing, amplification and other methodologies, and in looking for polymorphisms among the population in the database. The thus-identified polymorphism can then be entered into the database for each sample, and the database sorted (stratified) using that polymorphism as a sorting parameter to identify any patterns and 30 correlations that emerge, such as age correlated changes in the frequency of the identified marker. If a correlation is identified, the locus of the marker can be mapped and its function or effect assessed or deduced.

Thus, the databases here provide means for:

identification of significantly different allelic frequencies of genetic factors by comparing the occurrence or disappearance of the markers with increasing age in population and then associating the markers with a disease or a

5 biochemical pathway;

identification of significantly different allelic frequencies of disease causing genetic factors by comparing the male with the female population or comparing other selected stratified populations and associating the markers with a disease or a biochemical pathway;

10 identification of significantly different allelic frequencies of disease causing genetic factors by comparing different ethnic groups and associating the markers with a disease or a biochemical pathway that is known to occur in high frequency in the ethnic group;

profiling potentially functional variants of genes through the general  
15 panmixed population stratified according to age, sex, and ethnic origin and thereby demonstrating the contribution of the variant genes to the physical condition of the investigated population;

identification of functionally relevant gene variants by gene disequilibrium analysis performed within the general panmixed population stratified according  
20 to age, sex, and ethnic origin and thereby demonstrating their contribution to the physical condition of investigated population;

identification of potentially functional variants of chromosomes or parts of chromosomes by linkage disequilibrium analysis performed within the general panmixed population stratified according to age, sex, and ethnic origin and  
25 thereby demonstrating their contribution to the physical condition of investigated population.

#### **Uses of the identified markers and known markers**

The databases may also be used in conjunction with known markers and sorted to identify any correlations. For example, the databases can be used for:

30 determination and evaluation of the penetrance of medically relevant polymorphic markers;

determination and evaluation of the diagnostic specificity of medically relevant genetic factors;

determination and evaluation of the positive predictive value of medically relevant genetic factors;

- 5       determination and evaluation of the onset of complex diseases, such as, but are not limited to, diabetes, hypertension, autoimmune diseases, arteriosclerosis, cancer and other diseases within the general population with respect to their causative genetic factors;

delineation of the appropriate strategies for preventive disease treatment;

- 10       delineation of appropriate timelines for primary disease intervention;  
validation of medically relevant genetic factors identified in isolated populations regarding their general applicability;

validation of disease pathways including all potential target structures identified in isolated populations regarding their general applicability; and

- 15       validation of appropriate drug targets identified in isolated populations regarding their general applicability.

Among the diseases and disorders for which polymorphisms may be linked include, those linked to inborn errors of metabolism, acquired metabolic disorders, intermediary metabolism, oncogenesis pathways, blood clotting

- 20       pathways, and DNA synthetic and repair pathways DNA repair/replication/transcription factors and activities, *e.g.*, such as genes related to oncogenesis, aging and genes involved in blood clotting and the related biochemical pathways that are related to thrombosis, embolism, stroke, myocardial infarction, angiogenesis and oncogenesis.

- 25       For example, a number of diseases are caused by or involve deficient or defective enzymes in intermediary metabolism (see, *e.g.*, Tables 1 and 2, below) that result, upon ingestion of the enzyme substrates, in accumulation of harmful metabolites that damage organs and tissues, particularly an infant's developing brain and other organs, resulting in mental retardation and other developmental  
30       disorders.

Identification of markers and genes for such disorders is of great interest.

#### Model systems

Several gene systems, p21, p53 and Lipoprotein Lipase polymorphism (N291S), were selected. The p53 gene is a tumor suppressor gene that is  
5 mutated in diverse tumor types. One common allelic variant occurs at codon 72. A polymorphism that has been identified in the p53 gene, i.e., the R72P allele, results in an amino acid exchange, arginine to proline, at codon 72 of the gene.

Using diseased populations, it has been shown that there are ethnic differences in the allelic distribution of these alleles among African-Americans  
10 and Caucasians in the U.S. The results here support this finding and also demonstrate that the results obtained with a healthy database are meaningful (see, Figure 7B).

The 291S allele leads to reduced levels of high density lipoprotein cholesterol (HDL-C) that is associated with an increased risk of males for  
15 arteriosclerosis and in particular myocardial infarction (see, Reymer *et al.* (1995) *Nature Genetics* 10:28-34).

Both genetic polymorphisms were profiled within a part of the Caucasian population-based sample bank. For the polymorphism located in the lipoprotein lipase gene a total of 1025 unselected individuals (436 males and 589 females)  
20 were tested. Genomic DNA was isolated from blood samples obtained from the individuals.

As shown in the Examples and figures, an exemplary database containing about 5000 subjects, answers to the questionnaire (see Figure 3), and genotypic information has been stratified. A particular known allele has been selected, and  
25 the samples tested for the marker using mass spectrometric analyses, particularly PROBE (see the EXAMPLES) to identify polymorphisms in each sample. The population in the database has been sorted according to various parameters and correlations have been observed. For example, FIGURES 2A-C, show sorting of the data by age and sex for the Lipoprotein Lipase gene in the  
30 Caucasian population in the database. The results show a decrease in the frequency of the allele with age in males but no such decrease in females. Other

alleles that have been tested against the database, include, alleles of p53, p21 and factor VII. Results when sorted by age are shown in the figures.

These examples demonstrate an effect of altered frequency of disease causing genetic factors within the general population. The scientific

- 5 interpretation of those results allows prediction of medical relevance of polymorphic genetic alterations. In addition, conclusions can be drawn with regard to their penetrance, diagnostic specificity, positive predictive value, onset of disease, most appropriate onset of preventive strategies, and the general applicability of genetic alterations identified in isolated populations to panmixed
- 10 populations.

Therefore, an age- and sex-stratified population-based sample bank that is ethnically homogenous is a suitable tool for rapid identification and validation of genetic factors regarding their potential medical utility.

15 **Exemplary computer system for creating, storing and processing the databases Systems**

- Systems, including computers, containing the databases are provided herein. The computers and databases can be used in conjunction, for example, with the APL system (see, copending U.S. application Serial No. 09/285,481),
- 20 which is an automated system for analyzing biopolymers, particularly nucleic acids. Results from the APL system can be entered into the database.

- Any suitable computer system may be used. The computer system may be integrated into systems for sample analysis, such as the automated process line described herein (see, *e.g.*, copending U.S. application Serial No.
- 25 09/285,481).

- Figure 17 is a block diagram of a computer constructed in to provide and process the databases described herein. The processing that maintains the database and performs the methods and procedures may be performed on multiple computers all having a similar construction, or may be performed by a
- 30 single, integrated computer. For example, the computer through which data is added to the database may be separate from the computer through which the database is sorted, or may be integrated with it. In either arrangement, the

computers performing the processing may have a construction as illustrated in Figure 17.

Figure 17 is a block diagram of an exemplary computer 1700 that maintains the database described above and performs the methods and procedures. Each computer 1700 operates under control of a central processor unit (CPU) 1702, such as a "Pentium" microprocessor and associated integrated circuit chips, available from Intel Corporation of Santa Clara, California, USA. A computer user can input commands and data from a keyboard and display mouse 1704 and can view inputs and computer output at a display 1706. The display is typically a video monitor or flat panel display device. The computer 1700 also includes a direct access storage device (DASD) 1707, such as a fixed hard disk drive. The memory 1708 typically comprises volatile semiconductor random access memory (RAM). Each computer preferably includes a program product reader 1710 that accepts a program product storage device 1712, from which the program product reader can read data (and to which it can optionally write data). The program product reader can comprise, for example, a disk drive, and the program product storage device can comprise removable storage media such as a magnetic floppy disk, an optical CD-ROM disc, a CD-R disc, a CD-RW disc, or a DVD data disc. If desired, the computers can be connected so they can communicate with each other, and with other connected computers, over a network 1713. Each computer 1700 can communicate with the other connected computers over the network 1713 through a network interface 1714 that enables communication over a connection 1716 between the network and the computer.

The computer 1700 operates under control of programming steps that are temporarily stored in the memory 1708 in accordance with conventional computer construction. When the programming steps are executed by the CPU 1702, the pertinent system components perform their respective functions. Thus, the programming steps implement the functionality of the system as described above. The programming steps can be received from the DASD 1707, through the program product reader 1712, or through the network connection 1716. The storage drive 1710 can receive a program product, read

programming steps recorded thereon and transfer the programming steps into the memory 1708 for execution by the CPU 1702. As noted above, the program product storage device 1710 can comprise any one of multiple removable media having recorded computer-readable instructions, including  
5 magnetic floppy disks and CD-ROM storage discs. Other suitable program product storage devices can include magnetic tape and semiconductor memory chips. In this way, the processing steps necessary for operation can be embodied on a program product.

Alternatively, the program steps can be received into the operating  
10 memory 1708 over the network 1713. In the network method, the computer receives data including program steps into the memory 1708 through the network interface 1714 after network communication has been established over the network connection 1716 by well-known methods that will be understood by those skilled in the art without further explanation. The program steps are then  
15 executed by the CPU 1702 to implement the processing of the Garment Database system.

It should be understood that all of the computers of the system preferably have a construction similar to that shown in Figure 17, so that details described with respect to the Figure 17 computer 1700 will be understood to apply to all  
20 computers of the system 1700. This is indicated by multiple computers 1700 shown connected to the network 1713. Any one of the computers 1700 can have an alternative construction, so long as they can communicate with the other computers and support the functionality described herein.

Figure 18 is a flow diagram that illustrates the processing steps  
25 performed using the computer illustrated in Figure 17, to maintain and provide access to the databases, such as for identifying polymorphic genetic markers. In particular, the information contained in the database is stored in computers having a construction similar to that illustrated in Figure 17. The first step for maintaining the database, as indicated in Figure 18, is to identify healthy  
30 members of a population. As noted above, the population members are subjects that are selected only on the basis of being healthy, and where the subjects are mammals, such as humans, they are preferably selected based upon apparent

health and the absence of detectable infections. The step of identifying is represented by the flow diagram box numbered 1802.

The next step, represented by the flow diagram box numbered 1804, is to obtain identifying and historical information and data relating to the identified  
5 members of the population. The information and data comprise parameters for each of the population members, such as member age, ethnicity, sex, medical history, and ultimately genotypic information. Initially, the parameter information is obtained from a questionnaire answered by each member, from whom a body tissue or body fluid sample also is obtained. The step of entering and storing  
10 these parameters into the database of the computer is represented by the flow diagram box numbered 1806. As additional information about each population member and corresponding sample is obtained, this information can be inputted into the database and can serve as a sorting parameter.

In the next step, represented by the flow diagram box numbered 1808,  
15 the parameters of the members are associated with an indexer. This step may be executed as part of the database storage operation, such as when a new data record is stored according to the relational database structure and is automatically linked with other records according to that structure. The step  
1806 also may be executed as part of a conventional data sorting or retrieval  
20 process, in which the database entries are searched according to an input search or indexing key value to determine attributes of the data. For example, such search and sort techniques may be used to follow the occurrence of known genetic markers and then determine if there is a correlation with diseases for which they have been implicated. Examples of this use are for assessing the  
25 frequencies of the p53 and Lipoprotein Lipase polymorphisms.

Such searching of the database also may be valuable for identifying one or more genetic markers whose frequency changes within the population as a function of age, ethnic group, sex, or some other criteria. This can allow the identification of previously unknown polymorphisms and, ultimately,  
30 identification of a gene or pathway involved in the onset and progression of disease.

In addition, the database can be used for taking an identified polymorphism and ascertaining whether it changes in frequency when the data is sorted according to a selected parameter.

In this way, the databases and methods provided herein permit, among  
5 other things, identification of components, particularly key components, of a disease process by understanding its genetic underpinnings, and also an understanding of processes, such as individual drug responses. The databases and methods provided herein also can be used in methods involving elucidation of pathological pathways, in developing new diagnostic assays, identifying new  
10 potential drug targets, and in identifying new drug candidates.

**Morbidity and/or early mortality associated polymorphisms**

A database containing information provided by a population of healthy blood donors who were not selected for any particular disease to can be used to  
15 identify polymorphisms and the alleles in which they are present, whose frequency decreases with age. These may represent morbidity susceptibility markers and genes.

Polymorphisms of the genome can lead to altered gene function, protein function or genome instability. To identify those polymorphisms which have a clinical relevance/utility is the goal of a world-wide scientific effort. It can be  
20 expected that the discovery of such polymorphisms will have a fundamental impact on the identification and development of novel drug compounds to cure diseases. However, the strategy to identify valuable polymorphisms is cumbersome and dependent upon the availability of many large patient and control cohorts to show disease association. In particular, genes that cause a  
25 general risk of the population to suffer from any disease (morbidity susceptibility genes) will escape these case/control studies entirely.

Here described is a screening strategy to identify morbidity susceptibility genes underlying a variety of different diseases. The definition of a morbidity susceptibility gene is a gene that is expressed in many different cell types or  
30 tissues (housekeeping gene) and its altered function can facilitate the expression of a clinical phenotype caused by disease-specific susceptibility genes that are involved in a pathway specific for this disorder. In other words, these morbidity

susceptibility genes predispose people to develop a distinct disease according to their genetic make-up for this disease.

Candidates for morbidity susceptibility genes can be found at the bottom level of pathways involving transcription, translation, heat-shock proteins, protein trafficking, DNA repair, assembly systems for subcellular structures (e.g. mitochondria, peroxysomes and other cellular microbodies), receptor signaling cascades, immunology, etc. Those pathways control the quality of life at the cellular level as well as for the entire organism. Mutations/polymorphisms located in genes encoding proteins for those pathways can reduce the fitness of cells and make the organism more susceptible to express the clinical phenotype caused by the action of a disease-specific susceptibility gene. Therefore, these morbidity susceptibility genes can be potentially involved in a whole variety of different complex diseases if not in all. Disease-specific susceptibility genes are involved in pathways that can be considered as disease-specific pathways like glucose-, lipid, hormone metabolism, etc.

The exemplified method permit, among other things, identification of genes and/or gene products involved in a man's general susceptibility to morbidity and/or mortality; use of these genes and/or gene products in studies to elucidate the genetic underpinnings of human diseases; use of these genes and/or gene products in combinatorial statistical analyses without or together with disease-specific susceptibility genes; use of these genes and/or gene products to predict penetrance of disease susceptibility genes; use of these genes and/or gene products in predisposition and/or acute medical diagnostics and use of these genes and/or gene products to develop drugs to cure diseases and/or to extend the life span of humans.

#### SCREENING PROCESS

The healthy population stratified by age, gender and ethnicity, etc. is a very efficient and a universal screening tool for morbidity associated genes. Changes of allelic frequencies in the young compared to the old population are expected to indicate putative morbidity susceptibility genes. Individual samples of this healthy population base can be pooled to further increase the throughput. In a proof of principle experiment pools of young and old Caucasian females and

males were applied to screen more than 400 randomly chosen single nucleotide polymorphisms located in many different genes. Candidate polymorphisms were identified if the allelic difference was greater than 8% between young and old for both or only one of the genders. The initial results were assayed again in at least one independent subsequent experiments. Repeated experiments are necessary to recognize unstable biochemical reactions, which occur with a frequency of about 2-3% and can mimic age-related allelic frequency differences. Average frequency differences and standard deviations are calculated after successful reproducibility of initial results. The final allelic frequency is then compared to a reference population of Caucasian CEPH sample pool. The result should show similar allelic frequencies in the young Caucasian population. Subsequently, the exact allele frequencies of candidates including genotype information were obtained by analyzing all individual samples. This procedure is straight forward with regard to time and cost. It enables the screening of an enormous number of SNPs. So far, several markers with a highly significant association to age were identified and described below.

In general at least 5 individuals in a stratified population need to be screened to produce statistically significant results. The frequency of the allele is determined for an age stratified population. Chi square analysis is then performed on the allelic frequencies to determine if the difference between age groups is statistically significant. A p value less than of 0.1 is considered to represent a statistically significant difference. More preferably the p value should be less than 0.05.

#### **Clinical Trials**

The identification of markers whose frequency in a population decreases with age also allows for better designed and balanced clinical trials. Currently, if a clinical trial utilizes a marker as a significant endpoint in a study and the marker disappears with age, then the results of the study may be inaccurate. By using methods provided herein, it can be ascertained that if a marker decreases in frequency with age. This information can be considered and controlled when designing the study. For, example, an age independent marker could be substituted in its place.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

#### EXAMPLE 1

This example describes the use of a database containing information  
5 provided by a population of healthy blood donors who were not selected for any particular disease to determine the distribution of allelic frequencies of known genetic markers with age and by sex in a Caucasian subpopulation of the database. The results described in this example demonstrate that a disease-related genetic marker or polymorphism can be identified by sorting a healthy  
10 database by a parameter or parameters, such as age, sex and ethnicity.

##### Generating a database

Blood was obtained by venous puncture from human subjects who met blood bank criteria for donating blood. The blood samples were preserved with EDTA at pH 8.0 and labeled. Each donor provided information such as age, sex,  
15 ethnicity, medical history and family medical history. Each sample was labeled with a barcode representing identifying information. A database was generated by entering, for each donor, the subject identifier and information corresponding to that subject into the memory of a computer storage medium using commercially available software, e.g., Microsoft Access.

##### 20 Model genetic markers

The frequencies of polymorphisms known to be associated at some level with disease were determined in a subpopulation of the subjects represented in the database. These known polymorphisms occur in the p21, p53 and Lipoprotein Lipase genes. Specifically, the N291S polymorphism (N291S) of the  
25 Lipoprotein Lipase gene, which results in a substitution of a serine for an asparagine at amino acid codon 291, leads to reduced levels of high density lipoprotein cholesterol (HDL-C) that is associated with an increased risk of males for arteriosclerosis and in particular myocardial infarction (see, Reymer *et al.* (1995) *Nature Genetics* 10:28-34).

30 The p53 gene encodes a cell cycle control protein that assesses DNA damage and acts as a transcription factor regulating genes that control cell growth, DNA repair and apoptosis (programmed cell death). Mutations in the

p53 gene have been found in a wide variety of different cancers, including different types of leukemia, with varying frequency. The loss of normal p53 function results in genomic instability and uncontrolled cell growth. A polymorphism that has been identified in the p53 gene, i.e., the R72P allele, results in the substitution of a proline for an arginine at amino acid codon 72 of the gene.

The p21 gene encodes a cyclin-dependent kinase inhibitor associated with G1 phase arrest of normal cells. Expression of the p21 gene triggers apoptosis. Polymorphisms of the p21 gene have been associated with Wilms' tumor, a pediatric kidney cancer. One polymorphism of the p21 gene, the S31R polymorphism, results in a substitution of an arginine for a serine at amino acid codon 31.

#### Database analysis

##### Sorting of subjects according to specific parameters

The genetic polymorphisms were profiled within segments of the Caucasian subpopulation of the sample bank. For p53 profiling, the genomic DNA isolated from blood from a total of 1277 Caucasian subjects age 18-59 years and 457 Caucasian subjects age 60-79 years was analyzed. For p21 profiling, the genomic DNA isolated from blood from a total of 910 Caucasian subjects age 18-49 years and 824 Caucasian subjects age 50-79 years was analyzed. For lipoprotein lipase gene profiling, the genomic DNA from a total of 1464 Caucasian females and 1470 Caucasian males under 60 years of age and a total of 478 Caucasian females and 560 Caucasian males over 60 years of age was analyzed.

##### Isolation and analysis of genomic DNA

Genomic DNA was isolated from blood samples obtained from the individuals. Ten milliliters of whole blood from each individual was centrifuged at 2000 x g. One milliliter of the buffy coat was added to 9 ml of 155 mM  $\text{NH}_4\text{Cl}$ , 10 mM  $\text{KHCO}_3$ , and 0.1 mM  $\text{Na}_2\text{EDTA}$ , incubated 10 min at room temperature and centrifuged for 10 min at 2000 x g. The supernatant was removed, and the white cell pellet was washed in 155 mM  $\text{NH}_4\text{Cl}$ , 10 mM  $\text{KHCO}_3$  and 0.1 mM  $\text{Na}_2\text{EDTA}$  and resuspended in 4.5 ml of 50 mM Tris, 5 mM

EDTA and 1% SDS. Proteins were precipitated from the cell lysate by 6 mM ammonium acetate, pH 7.3, and then separated from the nucleic acids by centrifugation at 3000 x g. The nucleic acid was recovered from the supernatant by the addition of an equal volume of 100% isopropanol and  
5 centrifugation at 2000 x g. The dried nucleic acid pellet was hydrated in 10 mM Tris, pH 7.6, and 1 mM Na<sub>2</sub>EDTA and stored at 4° C.

Assays of the genomic DNA to determine the presence or absence of the known genetic markers were developed using the BiomassPROBE™ detection method (primer oligo base extension) reaction. This method uses a single  
10 detection primer followed by an oligonucleotide extension step to give products, which can be readily resolved by mass spectrometry, and, in particular, MALDI-TOF mass spectrometry. The products differ in length depending on the presence or absence of a polymorphism. In this method, a detection primer anneals adjacent to the site of a variable nucleotide or sequence of nucleotides  
15 and the primer is extended using a DNA polymerase in the presence of one or more dideoxynTPs and, optionally, one or more deoxyNTPs. The resulting products are resolved by MALDI-TOF mass spectrometry. The mass of the products as measured by MALDI-TOF mass spectrometry makes possible the determination of the nucleotide(s) present at the variable site.

20 First, each of the Caucasian genomic DNA samples was subjected to nucleic acid amplification using primers corresponding to sites 5' and 3' of the polymorphic sites of the p21 (S31R allele), p53 (R72P allele) and Lipoprotein Lipase (N291S allele) genes. One primer in each primer pair was biotinylated to permit immobilization of the amplification product to a solid support.  
25 Specifically, the polymerase chain reaction primers used for amplification of the relevant segments of the p21, p53 and lipoprotein lipase genes are shown below: US4p21c31-2F (SEQ ID NO: 9) and US5p21-2R (SEQ ID NO: 10) for p21 gene amplification; US4-p53-ex4-F (also shown as p53-ex4US4 (SEQ ID NO: 2)) and US5-p53/2-4R (also shown as US5P53/4R (SEQ ID NO: 3)) for p53 gene  
30 amplification; and US4-LPL-F2 (SEQ ID NO: 16) and US5-LPL-R2 (SEQ ID NO: 17) for lipoprotein lipase gene amplification.

Amplification of the respective DNA sequences was conducted according to standard protocols. For example, primers may be used in a concentration of 8 pmol. The reaction mixture (e.g., total volume 50  $\mu$ l) may contain

Taq-polymerase including 10x buffer and dTNPs. Cycling conditions for

- 5 polymerase chain reaction amplification may typically be initially 5 min. at 95°C, followed by 1 min. at 94°C, 45 sec at 53°C, and 30 sec at 72°C for 40 cycles with a final extension time of 5 min at 72°C. Amplification products may be purified by using Qiagen's PCR purification kit (No. 28106) according to manufacturer's instructions. The elution of the purified products from the
- 10 column can be done in 50  $\mu$ l TE-buffer (10mM Tris, 1 mM EDTA, pH 7.5).

The purified amplification products were immobilized via a biotin-avidin linkage to streptavidin-coated beads and the double-stranded DNA was denatured. A detection primer was then annealed to the immobilized DNA using conditions such as, for example, the following: 50  $\mu$ l annealing buffer (20 mM

15 Tris, 10 mM KCl, 10 mM  $(\text{NH}_4)_2\text{SO}_4$ , 2 mM  $\text{MgSO}_2$ , 1% Triton X-100, pH 8) at 50°C for 10 min, followed by washing of the beads three times with 200  $\mu$ l washing buffer (40 mM Tris, 1 mM EDTA, 50 mM NaCl, 0.1% Tween 20, pH 8.8) and once in 200  $\mu$ l TE buffer.

- The PROBE extension reaction was performed, for example, by using
- 20 some components of the DNA sequencing kit from USB (No. 70770) and dNTPs or ddNTPs from Pharmacia. An exemplary protocol could include a total reaction volume of 45  $\mu$ l, containing of 21  $\mu$ l water, 6  $\mu$ l Sequenase-buffer, 3  $\mu$ l 10 mM DTT solution, 4.5  $\mu$ l, 0.5 mM of three dNTPs, 4.5  $\mu$ l, 2 mM the missing one ddNTP, 5.5  $\mu$ l glycerol enzyme dilution buffer, 0.25  $\mu$ l Sequenase 2.0, and 0.25
- 25 pyrophosphatase. The reaction can then be pipetted on ice and incubated for 15 min at room temperature and for 5 min at 37°C. The beads may be washed three times with 200  $\mu$ l washing buffer and once with 60  $\mu$ l of a 70 mM  $\text{NH}_4$ -Citrate solution.

- The DNA was denatured to release the extended primers from the
- 30 immobilized template. Each of the resulting extension products was separately analyzed by MALDI-TOF mass spectrometry using 3-hydroxypicolinic acid (3-HPA) as matrix and a UV laser.

Specifically, the primers used in the PROBE reactions are as shown below: P21/31-3 (SEQ ID NO: 12) for PROBE analysis of the p21 polymorphic site; P53/72 (SEQ ID NO: 4) for PROBE analysis of the p53 polymorphic site; and LPL-2 for PROBE analysis of the lipoprotein lipase gene polymorphic site. In the PROBE analysis of the p21 polymorphic site, the extension reaction was performed using dideoxy-C. The products resulting from the reaction conducted on a "wild-type" allele template (wherein codon 31 encodes a serine) and from the reaction conducted on a polymorphic S31R allele template (wherein codon 31 encodes an arginine) are shown below and designated as P21/31-3 Ser (wt) (SEQ ID NO: 13) and P21/31-3 Arg (SEQ ID NO: 14), respectively. The masses for each product as can be measured by MALDI-TOF mass spectrometry are also provided (i.e., 4900.2 Da for the wild-type product and 5213.4 Da for the polymorphic product).

In the PROBE analysis of the p53 polymorphic site, the extension reaction was performed using dideoxy-C. The products resulting from the reaction conducted on a "wild-type" allele template (wherein codon 72 encodes an arginine) and from the reaction conducted on a polymorphic R72P allele template (wherein codon 72 encodes a proline) are shown below and designated as Cod72 G Arg (wt) and Cod72 C Pro, respectively. The masses for each product as can be measured by MALDI-TOF mass spectrometry are also provided (i.e., 5734.8 Da for the wild-type product and 5405.6 Da for the polymorphic product).

In the PROBE analysis of the lipoprotein lipase gene polymorphic site, the extension reaction was performed using a mixture of ddA and ddT. The products resulting from the reaction conducted on a "wild-type" allele template (wherein codon 291 encodes an asparagine) and from the reaction conducted on a polymorphic N291S allele template (wherein codon 291 encodes a serine) are shown below and designated as 291Asn and 291Ser, respectively. The masses for each product as can be measured by MALDI-TOF mass spectrometry are also provided (i.e., 6438.2 Da for the wild-type product and 6758.4 Da for the polymorphic product).

P53-1 (R72P)

PCR Product length: 407 bp (SEQ ID NO: 1)

US4-p53-ex4-F  
 ctg aggacctggt cctctgactg  
 5 ctcttttcac ccattctacag tcccccttgc cgtcccaagc aatggatgat ttgatgctgt  
 ccccggaaga tattgaacaa tggttcactg aagaccagg tccagatgaa gctcccagaa  
 P53/72 72R  
 tggcagagggc tgctccccgc gtggcccctg caccagcagc tcctacaccg gcggcccctg  
 c 72P  
 10 caccagcccc ctccctggccc ctgtcatctt ctgtcccttc ccagaaaacc taccagggca  
 gctacgggtt ccgtctgggc ttcttgcat ctgggacagc caagtctgtg acttgacagg  
 tcagttgccc tgaggggctg gcttccatga gacttcaa  
 US5-p53/2-4R

Primers (SEQ ID NOs: 2-4)

p53-ex4FUS4 ccc agt cac gac gtt gta aaa cgc tga gga cct ggt cct ctg ac  
 15 US5P53/4R agc gga taa caa ttt cac aca ggt tga agt ctc atg gaa gcc  
 P53/72 gcc aga ggc tgc tcc cc

#### Masses

Allele	Product Termination: ddC	SEQ #	Length	Mass
P53/72	gccagaggctgctcccc	5	17	5132.4
20 Cod72 G Arg (wt)	gccagaggctgctccccgc	6	19	5734.8
Cod72 C Pro	gccagaggctgctcccc	7	18	5405.6

Biotinylated US5 primer is used in the PCR amplification.

#### LPL-1 (N291S)

25 Amino acid exchange asparagine to serine at codon 291 of the lipoprotein lipase gene.

PCR Product length: 251 bp (SEQ ID NO: 15)

US4-LPL-F2 (SEQ ID NO: 16)

30 gcgtccatt catctcttca tcgactctct gttgaatgaa gaaaatccaa gtaaggccta  
 caggtgcagt tccaaggaag cctttgagaa agggctctgc ttgagttgta gaaagaaccg  
 LPL-2 291N  
 ctgcaacaat ctgggctatg aqatcaataa agtcagagcc aaaagaagca gcaaaatgta  
 g 291S  
 35 cctgaagact cgttctcaga tgccc  
 US4-LPL-R2

Primers (SEQ ID NOs: 16-18):

US4-LPL-F2 ccc agt cac gac gtt gta aaa cgg cgc tcc att cat ctc ttc

US5-LPL-R2 agc gga taa caa ttt cac aca ggg ggc atc tga gaa cga gtc

LPL-2 caa tct ggg cta tga gat ca

**Masses**

Allele	Product Termination: ddA, ddT	SEQ #	Length	Mass
LPL-2	caatctgggctatgagatca	19	20	6141
291 Asn	caatctgggctatgagatcaa	20	21	6438.2
5 291 Ser	caatctgggctatgagatcagt	21	22	6758.4

Biotinylated US5 primer is used in the PCR amplification.

**P21-1 (S31R)**

Amino acid exchange serine to arginine at codon 31 of the tumor

10 suppressor gene p21. Product length: 207 bp (SEQ ID NO: 8)

US4p21c31-2F

gtcc gtcagaaccc atgcggcagc  
p21/31-3 31S  
15 aaggcctgcc gccgcctctt cggcccagtg gacagcgaac agctgagccg cgactgtgat  
gcgctaattgg cgggctgcat ccaggaggcc cgtgagcgat ggaacttcga ctttgtcacc  
gagacaccac tggaggg a 31R  
US5p21-2R

**Primers (SEQ ID NOs: 9-11)**

20 US4p21c31-2F ccc agt cac gac qtt gta aaa cgg tcc gtc aga acc cat gcg g

US5p21-2R agc gga taa caa ttt cac aca ggc tcc agt ggt gtc tcg gtg ac

P21/31-3 cag cga gca gct gag

**Masses**

Allele	Product Termination: ddC	SEQ #	Length	Mass
25 p21/31-3	cagcgagcagctgag	12	15	4627
P21/31-3 Ser (wt)	cagcgagcagctgagc	13	16	4900.2
P21/31-3 Arg	cagcgagcagctgagac	14	17	5213.4

Biotinylated US5 primer is used in the PCR amplification.

30 Each of the Caucasian subject DNA samples was individually analyzed by MALDI-TOF mass spectrometry to determine the identity of the nucleotide at the polymorphic sites. The genotypic results of each assay can be entered into the database. The results were then sorted according to age and/or sex to determine the distribution of allelic  
35 frequencies by age and/or sex. As depicted in the Figures showing

histograms of the results, in each case, there was a differential distribution of the allelic frequencies of the genetic markers for the p21, p53 and lipoprotein lipase gene polymorphisms.

Figure 8 shows the results of the p21 genetic marker assays  
5 reveals a statistically significant decrease (from 13.3% to 9.2%) in the frequency of the heterozygous genotype (S31R) in Caucasians with age (18-49 years of age compared to 50-79 years of age). The frequencies of the homozygous (S31 and R31) genotypes for the two age groups are also shown, as are the overall frequencies of the S31 and R31 alleles in  
10 the two age groups (designated as \*S31 and \*R31, respectively in the Figure).

Figures 7A-C shows the results of the p53 genetic marker assays and reveals a statistically significant decrease (from 6.7% to 3.7%) in the frequency of the homozygous polymorphic genotype (P72) in Caucasians  
15 with age (18-59 years of age compared to 60-79 years of age). The frequencies of the homozygous "wild-type" genotype (R72) and the heterozygous genotype (R72P) for the two age groups are also shown, as are the overall frequencies of the R72 and P72 alleles in the two age groups (designated as \*R72 and \*P72, respectively in the Figure). These  
20 results are consistent with the observation that allele is not benign, as p53 regulates expression of a second protein, p21, which inhibits cyclin-dependent kinases (CDKs) needed to drive cells through the cell-cycle (a mutation in either gene can disrupt the cell cycle leading to increased cell division).

25 Figure 2C shows the results of the lipoprotein lipase gene genetic marker assays reveals a statistically significant decrease (from 1.97% to 0.54%) in the frequency of the polymorphic allele (S291) in Caucasian males with age (see also Reymer *et al.* (1995) *Nature Genetics* 10:28-34).

The frequencies of this allele in Caucasian females of different age groups are also shown.

## EXAMPLE 2

This example describes the use of MALDI-TOF mass spectrometry  
5 to analyze DNA samples of a number of subjects as individual samples  
and as pooled samples of multiple subjects to assess the presence or  
absence of a polymorphic allele (the 353Q allele) of the Factor VII gene  
and determine the frequency of the allele in the group of subjects. The  
results of this study show that essentially the same allelic frequency can  
10 be obtained by analyzing pooled DNA samples as by analyzing each  
sample separately and thereby demonstrate the quantitative nature of  
MALDI-TOF mass spectrometry in the analysis of nucleic acids.

### Factor VII

Factor VII is a serine protease involved in the extrinsic blood  
15 coagulation cascade. This factor is activated by thrombin and works with  
tissue factor (Factor III) in the processing of Factor X to Factor Xa. There  
is evidence that supports an association between polymorphisms in the  
Factor VII gene and increased Factor VII activity which can result in an  
elevated risk of ischemic cardiovascular disease, including myocardial  
20 infarction. The polymorphism investigated in this study is R353Q (i.e., a  
substitution of a glutamic acid residue for an arginine residue at codon  
353 of the Factor VII gene) (see Table 5).

### Analysis of DNA samples for the presence or absence of the 353Q allele of the Factor VII gene

25 Genomic DNA was isolated from separate blood samples obtained  
from a large number of subjects divided into multiple groups of 92  
subjects per group. Each sample of genomic DNA was analyzed using  
the BiomassPROBE™ assay as described in Example 1 to determine the  
30 presence or absence of the 353Q polymorphism of the Factor VII gene.

First, DNA from each sample was amplified in a polymerase chain reaction using primers F7-353FUS4 (SEQ ID NO: 24) and F7-353RUS5 (SEQ ID NO: 26) as shown below and using standard conditions, for example, as described in Example 1. One of the primers was biotinylated to permit immobilization of the amplification product to a solid support. The purified amplification products were immobilized via a biotin-avidin linkage to streptavidin-coated beads and the double-stranded DNA was denatured. A detection primer was then annealed to the immobilized DNA using conditions such as, for example, described in Example 1. The detection primer is shown as F7-353-P (SEQ ID NO: 27) below. The PROBE extension reaction was carried out using conditions, for example, such as those described in Example 1. The reaction was performed using ddG.

The DNA was denatured to release the extended primers from the immobilized template. Each of the resulting extension products was separately analyzed by MALDI-TOF mass spectrometry. A matrix such as 3-hydroxypicolinic acid (3-HPA) and a UV laser could be used in the MALDI-TOF mass spectrometric analysis. The products resulting from the reaction conducted on a "wild-type" allele template (wherein codon 353 encodes an arginine) and from the reaction conducted on a polymorphic 353Q allele template (wherein codon 353 encodes a glutamic acid) are shown below and designated as 353 CGG and 353 CAG, respectively. The masses for each product as can be measured by MALDI-TOF mass spectrometry are also provided (i.e., 5646.8 Da for the wild-type product and 5960 Da for the polymorphic product).

The MALDI-TOF mass spectrometric analyses of the PROBE reactions of each DNA sample were first conducted separately on each sample (250 nanograms total concentration of DNA per analysis). The allelic frequency of the 353Q polymorphism in the group of 92 subjects

was calculated based on the number of individual subjects in which it was detected.

Next, the samples from 92 subjects were pooled (250 nanograms total concentration of DNA in which the concentration of any individual DNA is 2.7 nanograms) and the pool of DNA was subjected to MALDI-TOF mass spectrometric analysis. The area under the signal corresponding to the mass of the 353Q polymorphism PROBE extension product in the resulting spectrum was integrated in order to quantitate the amount of DNA present. The ratio of this amount to total DNA was used to determine the allelic frequency of the 353Q polymorphism in the group of subjects. This type of individual sample vs. pooled sample analysis was repeated for numerous different groups of 92 different samples.

The frequencies calculated based on individual MALDI-TOF mass spectrometric analysis of the 92 separate samples of each group of 92 are compared to those calculated based on MALDI-TOF mass spectrometric analysis of pools of DNA from 92 samples in Figure 9. These comparisons are shown as "pairs" of bar graphs in the Figure, each pair being labeled as a separate "pool" number, e.g., P1, P16, P2, etc. Thus, for example, for P1, the allelic frequency of the polymorphism calculated by separate analysis of each of the 92 samples was 11.41% and the frequency calculated by analysis of a pool of all of the 92 DNA samples was 12.09%.

The similarity in frequencies calculated by analyzing separate DNA samples individually and by pooling the DNA samples demonstrates that it is possible, through the quantitative nature of MALDI-TOF mass spectrometry, to analyze pooled samples and obtain accurate frequency determinations. The ability to analyze pooled DNA samples significantly reduces the time and costs involved in the use of the non-selected, healthy databases as described herein. It has also been shown that it is

possible to decrease the DNA concentration of the individual samples in a pooled mixture from 2.7 nanograms to 0.27 nanograms without any change in the quality of the spectrum or the ability to quantitate the amount of sample detected.

## 5 Factor VII R353Q PROBE Assay

PROBE Assay for cod353 CGG>CAG (Arg>Gln), Exon 9 G>A.

PCR fragment: 134 bp (incl. US tags; SEQ ID Nos. 22 and 23)

Frequency of A allele: Europeans about 0.1, Japanese/Chinese about 0.03-0.05 (Thromb. Haemost. 1995, 73:617-22; Diabetologia 1998,

10 41:760-6):

### F7-353FUS4>

1201 GTGCCGGCTA CTCGGATGGC AGCAAGGACT CCTGCAAGGG GGACAGTGGG  
GGCCACATG

15 F7-353-P> A <F7-353RUS5  
1261 CCACCCACTA CCGGGGCACG TGGTACCTGA CGGGCATCGT CAGCTGGGGC  
CAGGGCTGCG

Primers (SEQ ID NOS: 24-26)

F7-353FUS4 CCC AGT CAC GAC GTT GTA AAA CGA TGG CAG CAA GGA CTC CTG Tm<sup>90</sup> 64°C

F7-353-P CAC ATG CCA CCC ACT ACC

20 F7-353RUS5 AGC GGA TAA CAA TTT CAC ACA GGT GAC GAT GCC CGT CAG GTA C 64°C

### Masses

Allele	Product Termination: ddG	SEQ #	Length	Mass
F7-353-P	atgccaccactacc	27	18	5333.6
353 CGG	cacatgccaccactaccg	28	19	5646.8
25 353 CAG	cacatgccaccactaccag	29	20	5960
US5-bio bio-	agcggataacaatttcacacagg	30	23	7648.6

### Conclusion

The above examples demonstrate an effect of altered frequency of  
30 disease causing genetic factors within the general population.

Interpretation of those results allows prediction of the medical relevance of polymorphic genetic alterations. In addition, conclusions can be drawn with regard to their penetrance, diagnostic specificity, positive predictive value, onset of disease, most appropriate onset of preventive strategies,

and the general applicability of genetic alterations identified in isolated populations to panmixed populations. Therefore, an age- and sex-stratified population-based sample bank that is ethnically homogenous is a suitable tool for rapid identification and validation of genetic factors regarding their potential medical utility.

### EXAMPLE 3

#### MORBIDITY AND MORTALITY MARKERS

##### Sample Band and Initial Screening

Healthy samples were obtained through the blood bank of San Bernardino, CA. Donors signed prior to the blood collection a consent form and agreed that their blood will be used in genetic studies with regard to human aging. All samples were anonymized. Tracking back of samples is not possible.

##### Isolation of DNA from blood samples of a healthy donor population

Blood is obtained from a donor by venous puncture and preserved with 1mM EDTA pH 8.0. Ten milliliters of whole blood from each donor was centrifuged at 2000x g. One milliliter of the buffy coat was added to 9 milliliters of 155mM  $\text{NH}_4\text{Cl}$ , 10mM  $\text{KHCO}_3$ , and 0.1mM  $\text{Na}_2\text{EDTA}$ , incubated 10 minutes at room temperature and centrifuged for 10 minutes at 2000x g. The supernatant was removed, and the white cell pellet was washed in 155mM  $\text{NH}_4\text{Cl}$ , 10mM  $\text{KHCO}_3$ , and 0.1mM  $\text{Na}_2\text{EDTA}$  and resuspended in 4.5 milliliters of 50mM Tris, 5mM EDTA, and 1% SDS. Proteins were precipitated from the cell lysate by 6M Ammonium Acetate, pH 7.3, and separated from the nucleic acid by centrifugation 3000x g. The nucleic acid was recovered from the supernatant by the addition of an equal volume of 100% isopropanol and centrifugation at 2000x g. The dried nucleic acid pellet was hydrated in 10mM Tris pH 7.6 and 1mM  $\text{Na}_2\text{EDTA}$  and stored at 4C.

In this study, samples were pooled as shown in Table 1. Both parents of the blood donors were of Caucasian origin.

Table 1

Pool ID	Sex	Age-range	# individuals
SP1	Female	18-39 years	276
SP2	Males	18-39 years	276
SP3	Females	60-69 years	184
SP4	Males	60-79 years	368

- 10 More than 400 SNPs were tested using all four pools. After one test run 34 assays were selected to be re-assayed at least once. Finally, 10 assays showed repeatedly differences in allele frequencies of several percent and, therefore, fulfilled the criteria to be tested using the individual samples. Average allele frequency and standard deviation is
- 15 tabulated in Table 2.

Table 2

Assay ID	SP1	SP1-STD	SP2	SP2-STD	SP3	SP3-STD	SP4	SP4-STD
47861	0.457	0.028	0.433	0.042	0.384	0.034	0.380	0.015
47751	0.276	0.007	0.403	0.006	0.428	0.052	0.400	0.097
48319	0.676	0.013	0.627	0.018	0.755	0.009	0.686	0.034
48070	0.581	0.034	0.617	0.045	0.561	n.a.	0.539	0.032
49807	0.504	0.034	0.422	0.020	0.477	0.030	0.556	0.005
49534	0.537	0.017	0.503	n.a.	0.623	0.023	0.535	0.009
49733	0.560	0.006	0.527	0.059	0.546	0.032	0.436	0.016
49947	0.754	0.008	0.763	0.047	0.736	0.052	0.689	0.025
50128	0.401	0.022	0.363	0.001	0.294	0.059	0.345	0.013

63306	0.697	0.012	0.674	0.013	0.712	0.017	0.719	0.005
-------	-------	-------	-------	-------	-------	-------	-------	-------

So far, 7 out of the 10 potential morbidity markers were fully analyzed. Additional information about genes in which these SNPs are  
 5 located was gathered through publicly databases like Genbank.

#### AKAPS

Candidate morbidity and mortality markers include housekeeping genes, such as genes involved in signal transduction. Among such genes are the A-kinase anchoring proteins (AKAPs) genes, which participate in  
 10 signal transduction pathways involving protein phosphorylation. Protein phosphorylation is an important mechanism for enzyme regulation and the transduction of extracellular signals across the cell membrane in eukaryotic cells. A wide variety of cellular substrates, including enzymes, membrane receptors, ion channels and transcription factors, can be  
 15 phosphorylated in response to extracellular signals that interact with cells. A key enzyme in the phosphorylation of cellular proteins in response to hormones and neurotransmitters is cyclic AMP (cAMP)-dependent protein kinase (PKA). Upon activation by cAMP, PKA thus mediates a variety of cellular responses to such extracellular signals. An array of PKA isozymes  
 20 are expressed in mammalian cells. The PKAs usually exist as inactive tetramers containing a regulatory (R) subunit dimer and two catalytic (C) subunits. Genes encoding three C subunits ( $C\alpha$ ,  $C\beta$  and  $C\gamma$ ) and four R subunits ( $R\text{I}\alpha$ ,  $R\text{I}\beta$ ,  $R\text{II}\alpha$  and  $R\text{II}\beta$ ) have been identified [see Takio *et al.* (1982) *Proc. Natl. Acad. Sci. U.S. A.* 79:2544-2548; Lee *et al.* (1983)  
 25 *Proc. Natl. Acad. Sci. U.S. A.* 80:3608-3612; Jahnsen *et al.* (1996) *J. Biol. Chem.* 261:12352-12361; Clegg *et al.* (1988) *Proc. Natl. Acad. Sci. U.S. A.* 85:3703-3707; and Scott (1991) *Pharmacol. Ther.* 50:123-145]. The type I (RI)  $\alpha$  and type II (RII)  $\alpha$  subunits are distributed ubiquitously, whereas  $R\text{I}\beta$  and  $R\text{II}\beta$  are present mainly in brain [see. *e.g.*, Miki and Eddy

(1999) *J. Biol. Chem.* 274:29057-29062]. The type I PKA holoenzyme (RI $\alpha$  and RI $\beta$ ) is predominantly cytoplasmic, whereas the majority of type II PKA (RII $\alpha$  and RII $\beta$ ) associates with cellular structures and organelles [Scott (1991) *Pharmacol. Ther.* 50:123-145]. Many hormones and other  
5 signals act through receptors to generate cAMP which binds to the R subunits of PKA and releases and activates the C subunits to phosphorylate proteins. Because protein kinases and their substrates are widely distributed throughout cells, there are mechanisms in place in cells to localize protein kinase-mediated responses to different signals. One  
10 such mechanism involves subcellular targeting of PKAs through association with anchoring proteins, referred to as A-kinase anchoring proteins (AKAPs), that place PKAs in close proximity to specific organelles or cytoskeletal components and particular substrates thereby providing for more specific PKA interactions and localized responses [see,  
15 *e.g.*, Scott *et al.* (1990) *J. Biol. Chem.* 265:21561-21566; Bregman *et al.* (1991) *J. Biol. Chem.* 266:7207-7213; and Miki and Eddy (1999) *J. Biol. Chem.* 274:29057-29062]. Anchoring not only places the kinase close to preferred substrates, but also positions the PKA holoenzyme at sites where it can optimally respond to fluctuations in the second messenger  
20 cAMP [Mochly-Rosen (1995) *Science* 268:247-251; Faux and Scott (1996) *Trends Biochem. Sci.* 21:312-315; Hubbard and Cohen (1993) *Trends Biochem. Sci.* 18:172-177].

Up to 75% of type II PKA is localized to various intracellular sites through association of the regulatory subunit (RII) with AKAPs [see, *e.g.*,  
25 Hausken *et al.* (1996) *J. Biol. Chem.* 271:29016-29022]. RII subunits of PKA bind to AKAPs with nanomolar affinity [Carr *et al.* (1992) *J. Biol. Chem.* 267:13376-13382], and many AKAP-RII complexes have been isolated from cell extracts. RI subunits of PKA bind to AKAPs with only micromolar affinity [Burton *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.*

94:11067-11072]. Evidence of binding of a PKA RI subunit to an AKAP has been reported [Miki and Eddy (1998) *J. Biol. Chem.* 273:34384-34390] in which RI $\alpha$ -specific and RI $\alpha$ /RII $\alpha$  dual specificity PKA anchoring domains were identified on FSC1/AKAP82. Additional dual specific  
5 AKAPs, referred to as D-AKAP1 and D-AKAP2, which interact with the type I and type II regulatory subunits of PKA have also been reported [Huang *et al.* (1997) *J. Biol. Chem.* 272:8057-8064; Huang *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:11184-11189].

More than 20 AKAPs have been reported in different tissues and  
10 species. Complementary DNAs (cDNAs) encoding AKAPs have been isolated from diverse species, ranging from *Caenorhabditis elegans* and *Drosophila* to human [see, *e.g.*, Colledge and Scott (1999) *Trends Cell Biol.* 9:216-221]. Regions within AKAPs that mediate association with RII subunits of PKA have been identified. These regions of approximately  
15 10-18 amino acid residues vary substantially in primary sequence, but secondary structure predictions indicate that they are likely to form an amphipathic helix with hydrophobic residues aligned along one face of the helix and charged residues along the other [Carr *et al.* (1991) *J. Biol. Chem.* 266:14188-14192; Carr *et al.* (1992) *J. Biol. Chem.* 267:13376-  
20 13382]. Hydrophobic amino acids with a long aliphatic side chain, *e.g.*, valine, leucine or isoleucine, may participate in binding to RII subunits [Glantz *et al.* (1993) *J. Biol. Chem.* 268:12796-12804].

Many AKAPs also have the ability to bind to multiple proteins, including other signaling enzymes. For example, AKAP79 binds to PKA,  
25 protein kinase C (PKC) and the protein phosphatase calcineurin (PP2B) [Coghlan *et al.* (1995) *Science* 267:108-112 and Klauck *et al.* (1996) *Science* 271:1589-1592]. Therefore, the targeting of AKAP79 to neuronal postsynaptic membranes brings together enzymes with opposite catalytic activities in a single complex.

AKAPs thus serve as potential regulatory mechanisms that increase the selectivity and intensity of a cAMP-mediated response. There is a need, therefore, to identify and elucidate the structural and functional properties of AKAPs in order to gain a complete understanding of the important role these proteins play in the basic functioning of cells.

#### **AKAP10**

The sequence of a human AKAP10 cDNA (also referred to as D-AKAP2) is available in the GenBank database, at accession numbers AF037439 (SEQ ID NO: 31) and NM 007202. The AKAP10 gene is located on chromosome 17.

The sequence of a mouse D-AKAP2 cDNA is also available in the GenBank database (see accession number AF021833). The mouse D-AKAP2 protein contains an RGS domain near the amino terminus that is characteristic of proteins that interact with  $G\alpha$  subunits and possess GTPase activating protein-like activity [Huang *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:11184-11189]. The human AKAP10 protein also has sequences homologous to RGS domains. The carboxy-terminal 40 residues of the mouse D-AKAP2 protein are responsible for the interaction with the regulatory subunits of PKA. This sequence is fairly well conserved between the mouse D-AKAP2 and human AKAP10 proteins.

#### **Polymorphisms of the human AKAP10 gene and polymorphic AKAP10 proteins**

Polymorphisms of AKAP genes that alter gene expression, regulation, protein structure and/or protein function are more likely to have a significant effect on the regulation of enzyme (particularly PKA) activity, cellular transduction of signals and responses thereto and on the basic functioning of cells than polymorphisms that do not alter gene and/or protein function. Included in the polymorphic AKAPs provided herein are human AKAP10 proteins containing differing amino acid residues at position number 646.

Amino acid 646 of the human AKAP10 protein is located in the carboxy-terminal region of the protein within a segment that participates in the binding of R-subunits of PKAs. This segment includes the carboxy-terminal 40 amino acids.

- 5       The amino acid residue reported for position 646 of the human AKAP10 protein is an isoleucine. Polymorphic human AKAP10 proteins provided herein have the amino acid sequence but contain residues other than isoleucine at amino acid position 646 of the protein. In particular  
10       embodiments of the polymorphic human AKAP10 proteins provided herein, the amino acid at position 646 is a valine, leucine or phenylalanine residue.

**An A to G transition at nucleotide 2073 of the human AKAP10 coding sequence**

- 15       As described herein, an allele of the human AKAP10 gene that contains a specific polymorphism at position 2073 of the coding sequence and thereby encodes a valine at position 646 has been detected in varying frequencies in DNA samples from younger and older segments of the human population. In this allele, the A at position 2073 of the AKAP10 gene coding sequence is changed from an A to a G, giving rise  
20       to an altered sequence in which the codon for amino acid 646 changes from ATT, coding for isoleucine, to GTT, coding for valine.

**Morbidity marker 1: human protein kinase A anchoring protein (AKAP10-1)**

- 25       PCR Amplification and BiomassPROBE assay detection of AKAP10-1 in a healthy donor population

PCR Amplification of donor population for AKAP 10

- PCR primers were synthesized by OPERON using phosphoramidite chemistry. Amplification of the AKAP10 target sequence was carried out in single 50 $\mu$ l PCR reaction with 100ng-1 $\mu$ g of pooled human genomic  
30       DNAs in a 50 $\mu$ l PCR reaction. Individual DNA concentrations within the

- pooled samples were present in equal concentration with the final concentration ranging from 1-25ng. Each reaction containing IX PCR buffer (Qiagen, Valencia, CA), 200uM dNTPs, 1U Hotstar Taq polymerase (Qiagen, Valencia, CA), 4mM MgCl<sub>2</sub>, and 25pmol of the
- 5 forward primer containing the universal primer sequence and the target specific sequence 5'-TCTCAATCATGTGCATTGAGG-3'(SEQ ID NO: 45), 2pmol of the reverse primer
- 5'-AGCGGATAACAATTTACACAGGGATCACACAGCCATCAGCAG-3' (SEQ ID NO: 46), and 10pmol of a biotinylated universal primer
- 10 complementary to the 5' end of the PCR amplicon
- 5'-AGCGGATAACAATTTACACAGG-3'(SEQ ID NO: 47). After an initial round of amplification with the target with the specific forward and reverse primer, the 5' biotinylated universal primer then hybridized and acted as a reverse primer thereby introducing a 3' biotin capture moiety
- 15 into the molecule. The amplification protocol results in a 5'-biotinylated double stranded DNA amplicon and dramatically reduces the cost of high throughput genotyping by eliminating the need to 5' biotin label each forward primer used in a genotyping. Thermal cycling was performed in 0.2mL tubes or 96 well plate using an MJ Research Thermal Cycler
- 20 (calculated temperature) with the following cycling parameters: 94° C for 5 min; 45 cycles: 94° C for 20 sec, 56° C for 30 sec, 72° C for 60 sec; 72° C 3min.

#### Immobilization of DNA

- The 50µl PCR reaction was added to 25ul of streptavidin coated magnetic
- 25 bead (Dyna) prewashed three times and resuspended in 1M NH<sub>4</sub>Cl, 0.06M NH<sub>4</sub>OH. The PCR amplicons were allowed to bind to the beads for 15 minutes at room temperature. The beads were then collected with a magnet and the supernatant containing unbound DNA was removed. The unbound strand was release from the double stranded amplicons by

incubation in 100mM NaOH and washing of the beads three times with 10mM Tris pH 8.0.

BiomassPROBE assay analysis of donor population for AKAP10-1 (clone 48319)

- 5            Genotyping using the BiomassPROBE assay methods was carried out by resuspending the DNA coated magnetic beads in 26mM Tris-HCl pH 9.5, 6.5 mM MgCl<sub>2</sub> and 50mM each of dTTP and 50mM each of ddCTP, ddATP, ddGTP, 2.5U of a thermostable DNA polymerase (Amersham) and 20pmol of a template specific oligonucleotide PROBE
- 10 primer 5'-CTGGCGCCCACGTGGTCAA-3' (SEQ ID NO: 48) (Operon). Primer extension occurs with three cycles of oligonucleotide primer hybridization and extension. The extension products were analyzed after denaturation from the template with 50mM NH<sub>4</sub>Cl and transfer of 150nL each sample to a silicon chip preloaded with 150nL of H3PA matrix
- 15 material. The sample material was allowed to crystallize and was analyzed by MALDI-TOF (Bruker, PerSeptive). The SNP that is present in AKAP10-1 is a T to C transversion at nucleotide number 156277 of the sequence of a genomic clone of the AKAP10 gene (GenBank Accession No. AC005730) (SEQ ID NO: 36). SEQ ID NO: 35: represents the
- 20 nucleotide sequence of human chromosome 17, which contains the genomic nucleotide sequence of the human AKAP10 gene, and SEQ ID NO: 36 represents the nucleotide sequence of human chromosome 17, which contains the genomic nucleotide sequence of the human AKAP10-1 allele. The mass of the primer used in the BioMass probe reaction was
- 25 5500.6 daltons. In the presence of the SNP, the primer is extended by the addition of ddC, which has a mass of 5773.8. The wildtype gene results in the addition of dT and ddG to the primer to produce an extension product having a mass of 6101 daltons.

The frequency of the SNP was measured in a population of age selected healthy individuals. Five hundred fifty-two (552) individuals between the ages of 18-39 years (276 females, 276 males) and 552 individuals between the ages of 60-79 (184 females between the ages of 60-69, 368 males between the age of 60-79) were tested for the presence of the polymorphism localized in the non-translated 3' region of AKAP 10. Differences in the frequency of this polymorphism with increasing age groups were observed among healthy individuals. Statistical analysis showed that the significance level for differences in the allelic frequency for alleles between the "younger" and the "older" populations was  $p = 0.0009$  and for genotypes was  $p = 0.003$ . Differences between age groups are significant. For the total population allele significance is  $p = 0.0009$ , and genotype significance is  $p = 0.003$ .

This marker led to the best significant result with regard to allele and genotype frequencies in the age-stratified population. Figure 19 shows the allele and genotype frequency in both genders as well as in the entire population. For latter the significance for alleles was  $p = 0.0009$  and for genotypes was  $p = 0.003$ . The young and old populations were in Hardy-Weinberg equilibrium. A preferential change of one particular genotype was not seen.

The polymorphism is localized in the non-translated 3'-region of the gene encoding the human protein kinase A anchoring protein (AKAP10). The gene is located on chromosome 17. Its structure includes 15 exons and 14 intervening sequences (introns). The encoded protein is responsible for the sub-cellular localization of the cAMP-dependent protein kinase and, therefore, plays a key role in the G-protein mediated receptor-signaling pathway (Huang et al. PNAS (1007) 94:11184-11189). Since its localization is outside the coding region, this polymorphism is most likely in linkage disequilibrium (LD) with other non-synonymous

polymorphisms that could cause amino acid substitutions and subsequently alter the function of the protein. Sequence comparison of different Genbank database entries concerning this gene revealed further six potential polymorphisms of which two are supposed to change the  
 5 respective amino acid (see Table 3).

Table 3

Exon	Codon	Nucleotides	Amino acid
3	100	GCT > GCC	Ala > Ala
4	177	AGT > GTG	Met > Val
8	424	GGG > GGC	Gly > Gly
10	524	CCG > CTG	Pro > Leu
12	591	GTG > GTC	Val > Val
12	599	CGC > CGA	Arg > Arg

15 **Morbidity marker 2: human protein kinase A anchoring protein (AKAP10-5)**

**Discovery of AKAP10-5 Allele (SEQ ID NO: 33)**

Genomic DNA was isolated from blood (as described above) of seventeen (17) individuals with a genotype CC at the AKAP10-1 gene  
 20 locus and a single heterozygous individual (CT) (as described). A target sequence in the AKAP10-1 gene which encodes the C-terminal PKA binding domain was amplified using the polymerase chain reaction. PCR primers were synthesized by OPERON using phosphoramidite chemistry. Amplification of the AKAP10-1 target sequence was carried out in  
 25 individual 50 $\mu$ l PCR reaction with 25ng of human genomic DNA templates. Each reaction containing 1 X PCR buffer (Qiagen, Valencia, CA), 200 $\mu$ M dNTPs, IU Hotstar Taq polymerase (Qiagen, Valencia, CA), 4mM MgCl<sub>2</sub>, 25pmol of the forward primer (Ex13F) containing the universal primer sequence and the target specific sequence 5'-TCC CAA  
 30 AGT GCT GGA ATT AC-3' (SEQ ID NO: 53), and 2pmol of the reverse

primer (Ex14R) 5'-GTC CAA TAT ATG CAA ACA GTT G-3' (SEQ ID NO: 54). Thermal cycling was performed in 0.2mL tubes or 96 well plate using an MJ Research Thermal Cycler (MJ Research, Waltham, MA) (calculated temperature) with the following cycling parameters: 94° C for 5 min; 45 cycles; 94° C for 20 sec, 56° C for 30 sec, 72° C for 60 sec; 72° C 3min. After amplification the amplicons were purified using a chromatography (Mo Bio Laboratories (Solana Beach, CA)).

The sequence of the 18 amplicons, representing the target region, was determined using a standard Sanger cycle sequencing method with 25nmol of the PCR amplicon, 3.2uM DNA sequencing primer 5'-CCC ACA GCA GTT AAT CCT TC-3'(SEQ ID NO: 55), and chain terminating dRhodamine labeled 2', 3' dideoxynucleotides (PE Biosystems, Foster City, CA) using the following cycling parameters: 96° C for 15 seconds; 25 cycles: 55° C for 15 seconds, 60° C for 4 minutes. The sequencing products precipitated by 0.3M NaOAc and ethanol. The precipitate was centrifuged and dried. The pellets were resuspended in deionized formamide and separated on a 5% polyacrylimide gel. The sequence was determined using the "Sequencher" software (Gene Codes, Ann Arbor, MI).

The sequence of all 17 of the amplicons, which are homozygous for the AKAP10-1 SNP of the amplicons, revealed a polymorphism at nucleotide position 152171 (numbering for GenBank Accession No. AC005730 for AKAP10 genomic clone (SEQ ID NO: 35)) with A replaced by G. This SNP can also be designated as located at nucleotide 2073 of a cDNA clone of the wildtype AKAP10 (GenBank Accession No. AF037439) (SEQ ID NO: 31). The amino acid sequence of the human AKAP10 protein is provided as SEQ ID NO: 32. This single nucleotide polymorphism was designated as AKAP10-5 (SEQ ID NO: 33) and resulted in a substitution of a valine for an isoleucine residue at amino

acid position 646 of the amino acid sequence of human AKAP10 (SEQ ID NO: 32).

**PCR Amplification and BiomassPROBE assay detection of AKAP10-5 in a healthy donor population**

5       The healthy population stratified by age is a very efficient and a universal screening tool for morbidity associated genes by allowing for the detection of changes of allelic frequencies in the young compared to the old population. Individual samples of this healthy population base can be pooled to further increase the throughput.

10       Healthy samples were obtained through the blood bank of San Bernardino, CA. Both parents of the blood donors were of Caucasian origin. Practically a healthy subject, when human, is defined as human donor who passes blood bank criteria to donate blood for eventual use in the general population. These criteria are as follows: free of detectable  
15       viral, bacterial, mycoplasma, and parasitic infections; not anemic; and then further selected based upon a questionnaire regarding history (see Figure 3). Thus, a healthy population represents an unbiased population of sufficient health to donate blood according to blood bank criteria, and not further selected for any disease state. Typically such individuals are  
20       not taking any medications.

      PCR primers were synthesized by OPERON using phosphoramidite chemistry. Amplification of the AKAP10 target sequence was carried out in a single 50 $\mu$ l PCR reaction with 100ng- 1 $\mu$ g of pooled human genomic DNAs in a 50 $\mu$ l PCR reaction. Individual DNA concentrations within the  
25       pooled samples were present in equal concentration with the final concentration ranging from 1-25ng. Each reaction contained 1X PCR buffer (Qiagen, Valencia, CA), 200 $\mu$ M dNTPs, 1U Hotstar Taq polymerase (Qiagen, Valencia, CA), 4mM MgCl<sub>2</sub>, and 25pmol of the forward primer containing the universal primer sequence and the target specific  
30       sequence 5'-AGCGGATAACAATTTACACAGGGAGCTAGCTTGGAAGAT

TGC-3' (SEQ ID NO: 41), 2pmol of the reverse primer 5'-GTCCAATATATGCAAACAGTTG-3' (SEQ ID NO: 54), and 10pmol of a biotinylated universal primer complementary to the 5' end of the PCR amplicon BIO:5'-AGCGGATAACAATTTACACAGG-3' (SEQ ID NO: 43).

- 5 After an initial round of amplification with the target with the specific forward and reverse primer, the 5' biotinylated universal primer can then be hybridized and acted as a forward primer thereby introducing a 5' biotin capture moiety into the molecule. The amplification protocol resulted in a 5'-biotinylated double stranded DNA amplicon and
- 10 dramatically reduced the cost of high throughput genotyping by eliminating the need to 5' biotin label every forward primer used in a genotyping.

- Thermal cycling was performed in 0.2mL tubes or 96 well plate using an MJ Research Thermal Cycler (calculated temperature) with the
- 15 following cycling parameters: 94° C for 5 min; 45 cycles: 94° C for 20 sec, 56° C for 30 sec; 72° C for 60 sec; 72° C 3min.

#### **Immobilization of DNA**

- The 50  $\mu$ L PCR reaction was added to 25 $\mu$ L of streptavidin coated magnetic beads (DynaL, Oslo, Norway), which were prewashed three
- 20 times and resuspended in 1M NH<sub>4</sub>Cl, 0.06M NH<sub>4</sub>OH. The 5' end of one strand of the double stranded PCR amplicons were allowed to bind to the beads for 15 minutes at room temperature. The beads were then collected with a magnet and the supernatant containing unbound DNA was removed. The hybridized but unbound strand was released from the
- 25 double stranded amplicons by incubation in 100mM NaOH and washing of the beads three times with 10mM Tris pH 8.0.

#### **Detection of AKAP10-5 using BiomassPROBE™ Assay**

BiomassPROBE™ assay of primer extension analysis (see, U.S. Patent No. 6,043,031) of donor population for AKAP 10-5 (SEQ ID NO:

33) was performed. Genotyping using these methods was carried out by resuspending the DNA coated magnetic beads in 26mM Tris-HCL pH 9.5, 6.5 mM MgCl<sub>2</sub>, 50mM dTTP, 50mM each of ddCTP, ddATP, ddGTP, 2.5U of a thermostable DNA polymerase (Ambersham), and 20pmol of a  
 5 template specific oligonucleotide PROBE primer  
 5'-ACTGAGCCTGCTGCATAA-3' (SEQ ID NO: 44) (Operon). Primer extension occurs with three cycles of oligonucleotide primer with hybridization and extension. The extension products were analyzed after denaturation from the template with 50 mM NH<sub>4</sub>Cl and transfer of 150 nL  
 10 of each sample to a silicon chip preloaded with 150 nl of H3PA matrix material. The sample material was allowed to crystallize and analyzed by MALDI-TOF (Bruker, PerSeptive). The primer has a mass of 5483.6 daltons. The SNP results in the additional of a ddC to the primer, giving a mass of 5756.8 daltons for the extended product. The wild type results in  
 15 the addition a T and ddG to the primer giving a mass of 6101 daltons.

The frequency of the SNP was measured in a population of age selected healthy individuals. Seven hundred thirteen (713) individuals under 40 years of age (360 females, 353 males) and 703 individuals over 60 years of age (322 females, 381 males) were tested for the presence of  
 20 the SNP, AKAP10-5 (SEQ ID NO: 33). Results are presented below in Table 1.

TABLE 1					
AKAP10-5 (2073V) frequency comparison in 2 age groups					
			<40	>60	delta G allele
25	Female	Alleles			
		*G	38.6	34.6	4.0
		*A	61.4	65.4	
		Genotypes			
		G	13.9	11.8	2.1
		GA	49.4	45.7	
		A	36.7	42.5	

5	Male	Alleles	*G	41.4	37.0	4.4
			*A	58.6	63.0	
		Genotypes	G	18.4	10.8	7.7
			GA	45.9	52.5	
			A	35.7	36.7	
10	Total	Alleles	*G	40.0	35.9	4.1
			*A	60.0	64.1	
		Genotypes	G	16.1	11.2	4.9
			GA	47.7	49.4	
			A	36.2	39.4	

Figure 20 graphically shows these results of allele and genotype distribution in the age and sex stratified Caucasian population.

**Morbidity marker 3: human methionine sulfoxide reductase A (msrA)**

The age-related allele and genotype frequency of this marker in both genders and the entire population is shown in Figure 21. The decrease of the homozygous CC genotype in the older male population is highly significant.

**Methionine sulfoxide reductase A (#63306)**

PCR Amplification and BiomassPROBE assay detection of the human methionine sulfoxide reductase A (h-msr-A) in a healthy donor population

PCR Amplification of donor population for h-msr-A

PCR primers were synthesized by OPERON using phosphoramidite chemistry. Amplification of the AKAP10 target sequence was carried out in single 50 $\mu$ l PCR reaction with 100ng-1 $\mu$ g of pooled human genomic DNA templates in a 50 $\mu$ l PCR reaction. Individual DNA concentrations within the pooled samples were present in an equal concentration with

- the final concentration ranging from 1-25ng. Each reaction containing 1 X PCR buffer (Qiagen, Valencia, CA), 200 $\mu$ M dNTPs, 1U Hotstar Taq polymerase (Qiagen, Valencia, CA), 4mM MgCl<sub>2</sub>, 25pmol of the forward primer containing the universal primer sequence and the target specific
- 5 sequence 5'-TTTCTCTGCACAGAGAGGC-3' (SEQ ID NO: 49), 2pmol of the reverse primer
- 5'-AGCGGATAACAATTTACACAGGGCTGAAATCCTTCGCTTTACC-3' (SEQ ID NO: 50), and 10pmol of a biotinylated universal primer complementary to the 5' end of the PCR amplicon
- 10 5'-AGCGGATAACAATTTACACAGG-3' (SEQ ID NO: 51). After an initial round of amplification of the target with the specific forward and reverse primers, the 5' biotinylated universal primer was then hybridized and acted as a reverse primer thereby introducing a 3' biotin capture moiety into the molecule. The amplification protocol results in a 5'-biotinylated
- 15 double stranded DNA amplicon and and dramatically reduces the cost of high throughput genotyping by eliminating the need to 5' biotin label each forward primer used in a genotyping. Thermal cycling was performed in 0.2mL tubes or 96 well plate using an MJ Research Thermal Cycler (calculated temperature) with the following cycling parameters: 94° C for
- 20 5 min; 45 cycles: 94° C for 20 sec, 56° C for 30 sec, 72° C for 60 sec; 72° C 3min.

#### Immobilization of DNA

- The 50 $\mu$ l PCR reaction was added to 25ul of streptavidin coated magnetic bead (Dynal) prewashed three times and resuspended in 1M
- 25 NH<sub>4</sub>Cl, 0.06M NH<sub>4</sub>OH. The PCR amplicons were allowed to bind to the beads for 15 minutes at room temperature. The beads were then collected with a magnet and the supernatant containing unbound DNA was removed. The unbound strand was release from the double stranded

amplicons by incubation in 100mM NaOH and washing of the beads three times with 10mM Tris pH 8.0.

**BiomassPROBE assay analysis of donor population for h-msr A**

- Genotyping using the BiomassPROBE assay methods was carried
- 5 out by resuspending the he DNA coated magnetic beads in 26mM Tris-HCl pH 9.5, 6.5 mM MgCl<sub>2</sub>, 50mM of dTTPs and 50mM each of ddCTP, ddATP, ddGTP, 2.5U of a thermostable DNA polymerase (Amersham), and 20pmol of a template specific oligonucleotide PROBE primer 5'-CTGAAAAGGGAGAGAAAG-3' (Operon) (SEQ ID NO: 52).
- 10 Primer extension occurs with three cycles of oligonucleotide primer with hybridization and extension. The extension products were analyzed after denaturation from the template with 50mM NH<sub>4</sub>Cl and transfer of 150nl each sample to a silicon chip preloaded with 150nl of H3PA matrix material. The sample material was allowed to crystallize and analyzed by
- 15 MALDI-TOF (Bruker, PerSeptive). The SNP is represented as a T to C tranversion in the sequence of two ESTs. The wild type is represented by having a T at position 128 of GenBank Accession No. AW 195104, which represents the nucleotide sequence of an EST which is a portion of the wild type human msrA gene (SEQ ID NO: 39 ). The SNP is presented
- 20 as a C at position 129 of GenBank Accession No. AW 874187, which represents the nucleotide sequence of an EST which is a portion of an allele of the human msrA gene (SEQ ID NO: 40 ).

- In a genomic sequence the SNP is represented as an A to G transversion. The primer utilized in the BioMass probe reaction had a
- 25 mass of 5654.8 daltons. In the presence of the SNP the primer is extended by the incorporation of a ddC and has a mass of 5928. In the presence of the wildtype the primer is extended by adding a dT and a DDC to produce a mass of 6232.1 daltons.

The frequency of the SNP was measured in a population of age selected healthy individuals. Five hundred fifty-two (552) individuals between the ages of 18-39 years (276 females, 276 males and 552 individuals between the age of 60-79 (184 females between the ages of 60-69, 368 males between the age of 60-79) were tested for the presence of the polymorphism localized in the nontranslated 3'-region of h-msr-A.

Genotype difference between male age group among healthy individuals is significant. For the male population allele significance is  $p = 0.0009$  and genotype significance is  $p = 0.003$ . The age-related allele and genotype frequency of this marker in both genders and the entire population is shown in Figure 21. The decrease of the homozygous CC genotype in the older male population is highly significant.

The polymorphism is localized in the non-translated 3'-region of the gene encoding the human methionine sulfoxide reductase (h-msrA). The exact localization is 451 base pairs downstream the stop codon (TAA). It is very likely that this SNP is in linkage disequilibrium (LD) with another polymorphism more upstream in the coding or promoter region; thus, it is not directly cause morbidity. The enzyme methionine sulfoxide reductase has been proposed to exhibit multiple biological functions. It may serve to repair oxidative protein damage but also play an important role in the regulation of proteins by activation or inactivation of their biological functions (Moskovitz et al. (1990) PNAS 95:14071-14075). It has also been shown that its activity is significantly reduced in brain tissues of Alzheimer patients (Gabbita et al., (1999) J. Neurochem 73:1660-1666). It is scientifically conceivable that proteins involved in the metabolism of reactive oxygen species are associated to disease.

## CONCLUSION

The use of the healthy population provides for the identification of morbidity markers. The identification of proteins involved in the G-protein coupled signaling transduction pathway or in the detoxification of oxidative stress can be considered as convincing results. Further confirmation and validation of other potential polymorphisms already identified *in silico* in the gene encoding the human protein kinase A anchoring protein could even provide stronger association to morbidity and demonstrate that this gene product is a suitable pharmaceutical or diagnostic target.

## EXAMPLE 4

### MALDI-TOF Mass Spectrometry Analysis

All of the products of the enzyme assays listed below were analyzed by MALDI-TOF mass spectrometry. A diluted matrix solution (0.15  $\mu$ L) containing of 10:1 3-hydroxypicolinic acid:ammonium citrate in 1:1 water:acetonitrile diluted 2.5-fold with water was pipetted onto a SpectroChip (Sequenom, Inc.) and was allowed to crystallize. Then, 0.15  $\mu$ L of sample was added. A linear PerSeptive Voyager DE mass spectrometer or Bruker Biflex MALDI-TOF mass spectrometer, operating in positive ion mode, was used for the measurements. The sample plates were kept at 18.2 kV for 400 nm after each UV laser shot (approximate 250 laser shots total), and then the target voltage was raised to 20 kV. The original spectra were digitized at 500 MHz.

## EXAMPLE 5

### Sample Conditioning

Where indicated in the examples below, the products of the enzymatic digestions were purified with ZipTips (Millipore, Bedford, MA). The ZipTips were pre-wetted with 10  $\mu$ L 50% acetonitrile and equilibrated 4 times with 10  $\mu$ L 0.1 M TEAAc. The oligonucleotide fragments were

- bound to the C18 in the ZipTip material by continuous aspiration and dispensation of each sample into the ZipTip. Each digested oligonucleotide was conditioned by washing with 10  $\mu$ L 0.1 M TEAAc, followed by 4 washing steps with 10  $\mu$ L H<sub>2</sub>O. DNA fragments were eluted from the
- 5 Ziptip with 7  $\mu$ L 50% acetonitrile.

Any method for condition the samples may be employed. Methods for conditioning, which generally is used to increase peak resolution, are well known (see, *e.g.*, International PCT application No. WO 98/20019).

#### EXAMPLE 6

##### 10 DNA Glycosylase-Mediated Sequence Analysis

DNA Glycosylases modifies DNA at each position that a specific nucleobase resides in the DNA, thereby producing abasic sites. In a subsequent reaction with another enzyme, a chemical, or heat, the phosphate backbone at each abasic site can be cleaved.

- 15 The glycosylase utilized in the following procedures was uracil-DNA glycosylase (UDG). Uracil bases were incorporated into DNA fragments in each position that a thymine base would normally occupy by amplifying a DNA target sequence in the presence of uracil. Each uracil substituted DNA amplicon was incubated with UDG, which cleaved each uracil base
- 20 in the amplicon, and was then subjected to conditions that effected backbone cleavage at each abasic site, which produced DNA fragments. DNA fragments were subjected to MALDI-TOF mass spectrometry analysis. Genetic variability in the target DNA was then assessed by analyzing mass spectra.

- 25 Glycosylases specific for nucleotide analogs or modified nucleotides, as described herein, can be substituted for UDG in the following procedures. The glycosylase methods described hereafter, in conjunction with phosphate backbone cleavage and MALDI, can be used to analyze DNA fragments for the purposes of SNP scanning, bacteria

typing, methylation analysis, microsatellite analysis, genotyping, and nucleotide sequencing and re-sequencing.

**A. Genotyping**

A glycosylase procedure was used to genotype the DNA sequence encoding UCP-2 (Uncoupling Protein 2). The sequence for UCP-2 is deposited in GenBank under accession number AF096289. The sequence variation genotyped in the following procedure was a cytosine (C-allele) to thymine (T-allele) variation at nucleotide position 4790, which results in a alanine to valine mutation at position 55 in the UCP-2 polypeptide.

DNA was amplified using a PCR procedure with a 50  $\mu$ L reaction volume containing of 5 pmol biotinylated primer having the sequence 5'-TGCTTATCCCTGTAGCTACCCTGTCTTGGCCTTGACAGATCCAA-3' (SEQ ID NO: 91), 15 pmol non-biotinylated primer having the sequence 5'-AGCGGATAACAATTTACACAGGCCATCACACCGCGGTACTG-3' (SEQ ID NO: 92), 200  $\mu$ M dATP, 200  $\mu$ M dCTP, 200  $\mu$ M dGTP, 600  $\mu$ M dUTP (to fully replace dTTP), 1.5 mM to 3 mM  $MgCl_2$ , 1 U of HotStarTaq polymerase, and 25 ng of CEPH DNA. Amplification was effected with 45 cycles at an annealing temperature of 56°C.

The amplification product was then immobilized onto a solid support by incubating 50  $\mu$ L of the amplification reaction with 5  $\mu$ L of prewashed Dynabeads for 20 minutes at room temperature. The supernatant was removed, and the beads were incubated with 50  $\mu$ L of 0.1 M NaOH for 5 minutes at room temperature to denature the double-stranded PCR product in such a fashion that single-stranded DNA was linked to the beads. The beads were then neutralized by three washes with 50  $\mu$ L 10 mM TrisHCl (pH 8). The beads were resuspended in 10  $\mu$ L of a 60mM TrisHCl/1mM EDTA (pH 7.9) solution, and 1 U uracil DNA glycosylase was added to the solution for 45 minutes at 37°C to remove uracil nucleotides present in the single-stranded DNA linked to the beads.

The beads were then washed two times with 25  $\mu$ L of 10 mM TrisHCl (pH 8) and once with 10  $\mu$ L of water. The biotinylated strands were then eluted from the beads with 12  $\mu$ L of 2 M  $\text{NH}_4\text{OH}$  at 60°C for 10 minutes. The backbone of the DNA was cleaved by incubating the samples for 10 min at 95°C (with a closed lid), and ammonia was evaporated from the samples by incubating the samples for 11 min at 80°C.

The cleavage fragments were then analyzed by MALDI-TOF mass spectrometry as described in Example 4. The T-allele generated a unique fragment of 3254 Daltons. The C-allele generated a unique fragment of 4788 Daltons. These fragments were distinguishable in mass spectra. Thus, the above-identified procedure was successfully utilized to genotype individuals heterozygous for the C-allele and T-allele in UCP-2.

**B. Glycosylase Analysis Utilizing Pooled DNA Samples**

The glycosylase assay was conducted using pooled samples to detect genetic variability at the UCP-2 locus. DNA of known genotype was pooled from eleven individuals and was diluted to a fixed concentration of 5 ng/ $\mu$ L. The procedure provided in Example 3A was followed using 2 pmol of forward primer having a sequence of 5'-CCCAGTCACGACGTTGTAAAACGTCTTGGCCTTGCAGATCCAAG-3' (SEQ ID NO: 93) and 15 pmol of reverse primer having the sequence 5'-AGCGGATAACAATTTACACAGGCCATCACACCGCGGTACTG-3' (SEQ ID NO: 94). In addition, 5 pmol of biotinylated primer having the sequence 5'-bioCCCAGTCACGACGTTGTAAAACG 3' (SEQ ID NO: 97) may be introduced to the PCR reaction after about two cycles. The fragments were analyzed via MALDI-TOF mass spectroscopy (Example 4). As determined in Example 3A, the T-allele, which generated a unique fragment of 3254 Daltons, could be distinguished in mass spectra from the C-allele, which generated a unique fragment of 4788 Daltons. Allelic frequency in the pooled samples was quantified by integrating the area

under each signal corresponding to an allelic fragment. Integration was accomplished by hand calculations using equations well known to those skilled in the art. In the pool of eleven samples, this procedure suggested that 40.9% of the individuals harbored the T allele and 59.09% of the  
5 individuals harbored the C allele.

**C. Glycosylase-Mediated Microsatellite Analysis**

A glycosylase procedure was utilized to identify microsatellites of the Bradykinin Receptor 2 (BKR-2) sequence. The sequence for BKR-2 is deposited in GenBank under accession number X86173. BKR-2 includes  
10 a SNP in the promoter region, which is a C to T variation, as well as a SNP in a repeated unit, which is a G to T variation. The procedure provided in Example 3A was utilized to identify the SNP in the promoter region, the SNP in the microsatellite repeat region, and the number of repeated units in the microsatellite region of BKR-2. Specifically, a  
15 forward PCR primer having the sequence 5'-CTCCAGCTGGGCAGGAGTGC-3' (SEQ ID NO: 95) and a reverse primer having the sequence 5'-CACTTCAGTCGCTCCCT-3' (SEQ ID NO: 96) were utilized to amplify BKR-2 DNA in the presence of uracil. The amplicon was fragmented by UDG followed by backbone cleavage. The  
20 cleavage fragments were analyzed by MALDI-TOF mass spectrometry as described in Example 4.

With regard to the SNP in the BKR-2 promoter region having a C to T variation, the C-allele generated a unique fragment having a mass of 7342.4 Daltons and the T-allele generated a unique fragment having a  
25 mass of 7053.2 Daltons. These fragments were distinguishable in mass spectra. Thus, the above-identified procedure was successfully utilized to genotype individuals heterozygous for the C-allele and T-allele in the promoter region of BKR-2.

With regard to the SNP in the BKR-2 repeat region having a G to T variation, the T-allele generated a unique fragment having a mass of 1784 Daltons, which was readily detected in a mass spectrum. Hence, the presence of the T-allele was indicative of the G to T sequence variation in the repeat region of BKR-2.

In addition, the number of repeat regions was distinguished between individuals having two repeat sequences and individuals having three repeat sequences in BKR-2. The DNA of these individuals did not harbor the G to T sequence variation in the repeat sequence as each repeat sequence contained a G at the SNP locus. The number of repeat regions was determined in individual samples by calculating the area under a signal corresponding to a unique DNA fragment having a mass of 2771.6 Daltons. This signal in spectra generated from individuals having two repeat regions had an area that was thirty-three percent less than the area under the same signal in spectra generated from individuals having three repeat regions. Thus, the procedures discussed above can be utilized to genotype individuals for the number of repeat sequences present in BKR-2.

#### **D. Bisulfite Treatment Coupled with Glycosylase Digestion**

Bisulfite treatment of genomic DNA can be utilized to analyze positions of methylated cytosine residues within the DNA. Treating nucleic acids with bisulfite deaminates cytosine residues to uracil residues, while methylated cytosine remains unmodified. Thus, by comparing the sequence of a PCR product generated from genomic DNA that is not treated with bisulfite with the sequence of a PCR product generated from genomic DNA that is treated with bisulfite, the degree of methylation in a nucleic acid as well as the positions where cytosine is methylated can be deduced.

Genomic DNA (2  $\mu$ g) was digested by incubation with 1  $\mu$ L of a restriction enzyme at 37°C for 2 hours. An aliquot of 3 M NaOH was added to yield a final concentration of 0.3M NaOH in the digestion solution. The reaction was incubated at 37°C for 15 minutes followed by  
5 treatment with 5.35M urea, 4.44M bisulfite, and 10mM hydroquinone, where the final concentration of hydroquinone is 0.5 mM.

The sample that was treated with bisulfite (sample A) was compared to the same digestion sample that had not undergone bisulfite treatment (sample B). After sample A was treated with bisulfite as  
10 described above, sample A and sample B were amplified by a standard PCR procedure. The PCR procedure included the step of overlaying each sample with mineral oil and then subjecting the sample to thermocycling (20 cycles of 15 minutes at 55°C followed by 30 seconds at 95°C). The PCR reaction contained four nucleotide bases, C, A, G, and U. The  
15 mineral oil was removed from each sample, and the PCR products were purified with glassmilk. Sodium iodide (3 volumes) and glassmilk (5  $\mu$ L) were added to samples A and B. The samples were then placed on ice for 8 minutes, washed with 420  $\mu$ L cold buffer, centrifuged for 10 seconds, and the supernatant fractions were removed. This process was  
20 repeated twice and then 25  $\mu$ L of water was added. Samples were incubated for 5 minutes at 37 °C, were centrifuged for 20 seconds, and the supernatant fraction was collected, and then this incubation/centrifugation/supernatant fraction collection procedure was repeated. 50  $\mu$ L 0.1 M NaOH was then added to the samples to denature  
25 the DNA. The samples were incubated at room temperature for 5 minutes, washed three times with 50  $\mu$ L of 10 mM TrisHCl (pH 8), and resuspended in 10  $\mu$ L 60mM TrisHCl/1mM EDTA, pH 7.9.

The sequence of PCR products from sample A and sample B were then treated with 2U of UDG (MBI Fermentas) and then subjected to

backbone cleavage, as described herein. The resulting fragments from each of sample A and sample B were analyzed by MALDI-TOF mass spectroscopy as described in Example 4. Sample A gave rise to a greater number of fragments than the number of fragments arising from sample B, indicative that the nucleic acid harbored at least one methylated cytosine moiety.

### EXAMPLE 7

#### Fen-Ligase-Mediated Haplotyping

Haplotyping procedures permit the selection of a fragment from one of an individual's two homologous chromosomes and to genotype linked SNPs on that fragment. The direct resolution of haplotypes can yield increased information content, improving the diagnosis of any linked disease genes or identifying linkages associated with those diseases. In previous studies, haplotypes were typically reconstructed indirectly through pedigree analysis (in cases where pedigrees were available) through laborious and unreliable allele-specific PCR or through single-molecule dilution methods well known in the art.

A haplotyping procedure was used to determine the presence of two SNPs, referred to as SNP1 and SNP2, located on one strand in a DNA sample. The haplotyping procedure used in this assay utilized Fen-1, a site-specific "flap" endonuclease that cleaves DNA "flaps" created by the overlap of two oligonucleotides hybridized to a target DNA strand. The two overlapping oligonucleotides in this example were short arm and long arm allele-specific adaptors. The target DNA was an amplified nucleic acid that had been denatured and contained SNP1 and SNP2.

The short arm adaptor included a unique sequence not found in the target DNA. The 3' distal nucleotide of the short arm adaptor was identical to one of the SNP1 alleles. Moreover, the long arm adaptor included two regions: a 3' region complementary to the short arm and a

5' gene-specific region complementary to the fragment of interest adjacent to the SNP. If there was a match between the adaptor and one of the homologues, the Fen enzyme recognized and cleaved the overlapping flap. The short arm of the adaptor was then ligated to the remainder of the target fragment (minus the SNP site). This ligated fragment was used as the forward primer for a second PCR reaction in which only the ligated homologue was amplified. The second PCR product (PCR2) was then analyzed by mass spectrometry. If there was no match between the adaptors and the target DNA, there was no overlap, no cleavage by Fen-1, and thus no PCR2 product of interest.

If there was more than one SNP in the sequence of interest, the second SNP (SNP2) was found by using an adaptor that was specific for SNP2 and hybridizing the adaptor to the PCR2 product containing the first SNP. The Fen-ligase and amplification procedures were repeated for the PCR2 product containing the first SNP. If the amplified product yielded a second SNP, then SNP1 and SNP2 were on the same fragment.

If the SNP is unknown, then four allele-specific adaptors (e.g. C, G, A, and T) can be used to hybridize with the target DNA. The substrates are then treated with the Fen-ligase protocol, including amplification. The PCR2 products may be analyzed by PROBE, as described herein, to determine which adaptors were hybridized to the DNA target and thus identify the SNPs in the sequence.

A Fen-ligase assay was used to detect two SNPs present in Factor VII. These SNPs are located 814 base pairs apart from each other. SNP1 was located at position 8401 (C to T), and SNP2 was located at 9215 (G to A) (SEQ ID #).

#### A. First Amplification Step

A PCR product (PCR1) was generated for a known heterozygous individual at SNP1, a short distance from the 5' end of the SNP.

Specifically, a 10  $\mu$ L PCR reaction was performed by mixing 1.5 mM  $MgCl_2$ , 200  $\mu$ M of each dNTP, 0.5 U HotStar polymerase, 0.1  $\mu$ M of a forward primer having the sequence 5'-GCG CTC CTG TCG GTG CCA (SEQ ID NO: 56), 0.1  $\mu$ M of a reverse primer having the sequence 5'-GCC  
5 TGA CTG GTG GGG CCC (SEQ ID NO: 57), and 1 ng of genomic DNA. The annealing temperature was 58°C, and the amplification process yielded fragments that were 861 bp in length.

The PCR1 reaction mixture was divided in half and was treated with an exonuclease 1/SAP mixture (0.22  $\mu$ L mixture/5  $\mu$ L PCR1 reaction)  
10 which contained 1.0  $\mu$ L SAP and 0.1  $\mu$ L exon1. The exonuclease treatment was done for 30 minutes at 37°C and then 20 minutes at 85°C to denature the DNA.

#### B. Adaptor Oligonucleotides

A solution of allele-specific adaptors (C and T), containing of one  
15 long and one short oligonucleotide per adaptor, was prepared. The long arm and short arm oligonucleotides of each adaptor (10  $\mu$ M) were mixed in a 1:1 ratio and heated for 30 seconds at 95°C. The temperature was reduced in 2°C increments to 37°C for annealing. The C-adaptor had a short arm sequence of 5'-CAT GCA TGC ACG GTC (SEQ ID NO: 58) and  
20 a long arm sequence of 5'-CAG AGA GTA CCC CTC GAC CGT GCA TGC ATG (SEQ ID NO: 59). Hence, the long arm of the adaptor was 30 bp (15 bp gene-specific), and the short arm was 15bp. The T-adaptor had a short arm sequence of 5'-CAT GCA TGC ACG GTT (SEQ ID NO: 60) and a long arm sequence of 5'-GTA CGT ACG TGC CAA CTC CCC ATG AGA  
25 GAC (SEQ ID NO: 61). The adaptor could also have a hairpin structure in which the short and long arm are separated by a loop containing of 3 to 10 nucleotides (SEQ ID NO: 118).

#### C. FEN-ligase reaction

In two tubes (one tube for each allele-specific adaptor per sample) was placed a solution (Solution A) containing of 3.5  $\mu$ l 10 mM 16%PEG/50 mM MOPS, 1.2  $\mu$ l 25 mM  $MgCl_2$ , 1.5  $\mu$ l 10X Ampligase Buffer, and 2.5  $\mu$ l PCR1. Each tube containing Solution A was incubated  
5 at 95°C for 5 minutes to denature the PCR1 product. A second solution (Solution B) containing of 1.65  $\mu$ l Ampligase (Thermostable ligase, Epicentre Technologies), 1.65  $\mu$ l 200ng/ $\mu$ l MFEN (from *Methanococcus jannaschii*), and 3.0  $\mu$ l of an allele specific adaptor (C or T) was prepared. Thus, different variations of Solution B, each variation containing of  
10 different allele-specific adaptors, were made. Solution B was added to Solution A at 95°C and incubated at 55°C for 3 hours. The total reaction volume was 15.0  $\mu$ l per adaptor-specific reaction. For a bi-allelic system, 2 x 15.0  $\mu$ l reactions were required.

The Fen-ligase reaction in each tube was then deactivated by  
15 adding 8.0  $\mu$ l 10 mM EDTA. Then, 1.0  $\mu$ l exoIII/Buffer (70%/30%) solution was added to each sample and incubated 30 minutes at 37°C, 20 minutes at 70°C (to deactivate exoIII), and 5 minutes at 95°C (to denature the sample and dissociate unused adaptor from template). The samples were cooled in an ice slurry and purified on UltraClean PCR  
20 Clean-up (MoBio) spin columns which removed all fragments less than 100 base pairs in length. The fragments were eluted with 50  $\mu$ l  $H_2O$ .

#### D. Second Amplification Step

A second amplification reaction (PCR2) was conducted in each sample tube using the short arm adaptor (C or T) sequence as the forward  
25 primer (minus the SNP1 site). Only the ligated homologue was amplified. A standard PCR reaction was conducted with a total volume of 10.0  $\mu$ l containing of 1X Buffer (final concentration), 1.5 mM final concentration  $MgCl_2$ , 200  $\mu$ M final concentration dNTPs, 0.5 U HotStar polymerase, 0.1  $\mu$ M final concentration forward primer 5'-CAT GCA TGC ACG GT (SEQ ID

NO: 62), 0.1  $\mu$ M final concentration reverse primer 5'-GCC TGA CTG GTG GGG CCC (SEQ ID NO: 63), and 1.0  $\mu$ l of the purified FEN-ligase reaction solution. The annealing temperature was 58°C. The PCR2 product was analyzed by MALDI TOF mass spectroscopy as described in Example 4.

- 5 The mass spectrum of Fen SNP1 showed a mass of 6084.08 Daltons, representing the C allele.

**E. Genotyping Additional SNPs**

- The second SNP (SNP2) can be found by using an adaptor that is specific for SNP2 and hybridizing that adaptor to the PCR2 product
- 10 containing the first SNP. The Fen-ligase and amplification procedures are repeated for the PCR2 product containing the first SNP. If the amplified product yields a second SNP, then SN1 and SN2 are on the same fragment. The mass spectrum of SNP2, representing the T allele, showed a mass of 6359.88 Daltons.

- 15 This assay can also be performed upon pooled DNA to yield haplotype frequencies as described herein. The Fen-ligase assay can be used to analyze multiplexes as described herein.

**EXAMPLE 8**

**Nickase-Mediated Sequence Analysis**

- 20 A DNA nickase, or DNase, was used to recognize and cleave one strand of a DNA duplex. Two nickases used were NY2A nickase and NYS1 nickase (Megabase) which cleave DNA at the following sites:

NY2A: 5'...R AG...3'

3'...Y↓TC...5' where R = A or G and Y = C or T

- 25 NYS1: 5'...↓CC[A/G/T]...3'

3'... GG[T/C/A]...5'.

**A. Nickase Digestion**

- Tris-HCl (10 mM), KCl (10 mM, pH 8.3), magnesium acetate (25 mM), BSA (1 mg/mL), and 6 U of Cvi NY2A or Cvi NYS1 Nickase (Megabase Research) were added to 25 pmol of double-stranded
- 5 oligonucleotide template having a sequence of 5'-CGC AGG GTT TCC TCG TCG CAC TGG GCA TGT G-3' (SEQ ID NO: 90, Operon, Alameda, CA) synthesized using standard phosphoramidite chemistry. With a total volume of 20  $\mu$ L, the reaction mixture was incubated at 37°C for 5 hours, and the digestion products were purified using ZipTips (Millipore, Bedford,
- 10 MA) as described in Example 5. The samples were analyzed by MALDI-TOF mass spectroscopy as described in Example 1. The nickase Cvi NY2A yielded three fragments with masses 4049.76 Daltons, 5473.14 Daltons, and 9540.71 Daltons. The Cvi NYS1 nickase yielded fragments with masses 2063.18 Daltons, 3056.48 Daltons, 6492.81 Daltons, and
- 15 7450.14 Daltons.

**B. Nickase Digestion of Pooled Samples**

- DQA (HLA ClassII-DQ Alpha, expected fragment size = 225bp) was amplified from the genomic DNA of 100 healthy individuals. DQA was amplified using standard PCR chemistry in a reaction having a total
- 20 volume of 50  $\mu$ L containing of 10 mM Tris-HCl, 10 mM KCl (pH 8.3), 2.5 mM MgCl<sub>2</sub>, 200  $\mu$ M of each dNTP, 10 pmol of a forward primer having the sequence 5'-GTG CTG CAG GTG TAA ACT TGT ACC AG-3' (SEQ ID NO: 64), 10 pmol of a reverse primer having the sequence 5'-CAC GGA TCC GGT AGC AGC GGT AGA GTT G-3' (SEQ ID NO: 65), 1 U DNA
- 25 polymerase (Stoffel fragment, Perkin Elmer), and 200ng human genomic DNA (2ng DNA/individual). The template was denatured at 94°C for 5 minutes. Thermal cycling was continued with a touch-down program that included 45 cycles of 20 seconds at 94°C, 30 seconds at 56°C, 1

minute at 72°C, and a final extension of 3 minutes at 72°C. The crude PCR product was used in the subsequent nickase reaction.

The unpurified PCR product was subjected to nickase digestion. Tris-HCl (10 mM), KCl (10 mM, pH 8.3), magnesium acetate (25mM),  
5 BSA (1 mg/mL), and 5 U of Cvi NY2A or Cvi NYS1 Nickase (Megabase Research) were added to 25 pmol of the amplified template with a total reaction volume of 20 $\mu$ L. The mixture was then incubated at 37°C for 5 hours. The digestion products were purified with either ZipTips (Millipore, Bedford, MA) as described in Example 5. The samples were analyzed by  
10 MALDI-TOF mass spectroscopy as described in Example 4. This assay can also be used to do multiplexing and standardless genotyping as described herein.

To simplify the nickase mass spectrum, the two complementary strands can be separated after digestion by using a single-stranded  
15 undigested PCR product as a capture probe. This probe (preparation shown below in Example 8C) can be hybridized to the nickase fragments in hybridization buffer containing 200 mM sodium citrate and 1% blocking reagent (Boehringer Mannheim). The reaction is heated to 95°C for 5 minutes and cooled to room temperature over 30 minutes by using a  
20 thermal cycler (PTC-200 DNA engine, MJ Research, Waltham, MA). The capture probe-nickase fragment is immobilized on 140  $\mu$ g of streptavidin-coated magnetic beads. The beads are subsequently washed three times with 70 mM ammonium citrate. The captured single-stranded nickase fragments are eluted by heating to 80°C for 5 minutes in 5  $\mu$ L of 50 mM  
25 ammonium hydroxide.

### C. Preparation of Capture Probe

The capture probe is prepared by amplifying the human  $\beta$ -globin gene (3' end of intron 1 to 5' end of exon 2) via PCR methods in a total volume of 50  $\mu$ L containing of GeneAmp 1XPCR Buffer II, 10 mM Tris-

HCl, pH 8.3, 50 mM KCl, 2 mM MgCl<sub>2</sub>, 0.2 mM dNTP mix, 10pmol of each primer (forward primer 5'-ACTGGGCATGTGGAGACAG-3' (SEQ ID NO: 66) and biotinylated reverse primer bio5'-GCACTTTCTTGCCATGAG-3' (SEQ ID: 67), 2 U of AmpliTaq Gold, and 200 ng of human genomic DNA. The template is denatured at 94°C for 8 minutes. Thermal cycling is continued with a touch-down program that included 11 cycles of 20 seconds at 94°C, 30 seconds at 64°C, 1 minute at 72°C; and a final extension of 5 minutes at 72°C. The amplicon is purified using UltraClean™ PCR clean-up kit (MO Bio Laboratories, Solano Beach, CA).

10

### EXAMPLE 9

#### Multiplex Type IIS SNP Assay

A Type IIS assay was used to identify human gene sequences with known SNPs. The Type IIS enzyme used in this assay was Fok I which effected double-stranded cleavage of the target DNA. The assay involved the steps of amplification and Fok I treatment of the amplicon. In the amplification step, the primers were designed so that each PCR product of a designated gene target was less than 100 bases such that a Fok I recognition sequence was incorporated at the 5' and 3' end of the amplicon. Therefore, the fragments that were cleaved by Fok I included a center fragment containing the SNP of interest.

Ten human gene targets with known SNPs were analyzed by this assay. Sequences of the ten gene targets, as well as the primers used to amplify the target regions, are found in Table 5. The ten targets were lipoprotein lipase, prothrombin, factor V, cholesterol ester transfer protein (CETP), factor VII, factor XIII, HLA-H exon 2, HLA-H exon 4, methylenetetrahydrofolate reductase (MTHR), and P53 exon 4 codon 72.

Amplification of the ten human gene sequences were carried out in a single 50 µL volume PCR reaction with 20 ng of human genomic DNA

template in 5 PCR reaction tubes. Each reaction vial contained 1X PCR buffer (Qiagen), 200 $\mu$ M dNTPs, 1U Hotstar Taq polymerase (Qiagen), 4 mM MgCl<sub>2</sub>, and 10pmol of each primer. US8, having sequence of 5'TCAGTCACGACGTT3'(SEQ ID NO: 68), and US9, having sequence of 5'CGGATAACAATTTC3'(SEQ ID NO: 69), were used for the forward and reverse primers respectively. Moreover, the primers were designed such that a Fok I recognition site was incorporated at the 5' and 3' ends of the amplicon. Thermal cycling was performed in 0.2 mL tubes or a 96 well plate using a MJ Research Thermal Cycler (calculated temperature) with the following cycling parameters: 94°C for 5 minutes; 45 cycles: 94°C for 20 seconds, 56°C for 20 seconds, 72°C for 60 seconds; and 72°C for 3 minutes.

Following PCR, the sample was treated with 0.2 U Exonuclease I (Amersham Pharmacia) and S Alkaline Phosphatase (Amersham Pharmacia) to remove the unincorporated primers and dNTPs. Typically, 0.2 U of exonuclease I and SAP were added to 5  $\mu$ L of the PCR sample. The sample was then incubated at 37°C for 15 minutes. Exonuclease I and SAP were then inactivated by heating the sample up to 85°C for 15 minutes. Fok I digestion was performed by adding 2 U of Fok I (New England Biolab) to the 5 uL PCR sample and incubating at 37°C for 30 minutes. Since the Fok I restriction sites are located on both sides of the amplicon, the 5' and 3' cutoff fragments have higher masses than the center fragment containing the SNP. The sample was then purified by anion exchange and analyzed by MALDI-TOF mass spectrometry as described in Example 4. The masses of the gene fragments from this multiplexing experiment are listed in Table 6. These gene fragments were resolved in mass spectra thereby allowing multiplex analysis of sequence variability in these genes.

**Table 5**  
**Genes for Multiplex Type IIS Assay**

5

10

15

Gene	Sequence	Seq. ID No.	Primers	Seq. ID No.
Lipoprotein Lipase (Asn291Ser)	cctttgagaa agggctctgc ttgagttgta gaaagaaccg ctgcaacaat <u>ctgggctatg agatca[a&gt;g]taa agtcagagcc</u> <u>aaaaaagaca gcaaaatgta</u>	98-99	5' caatttcacgctggatgcaatct gggctatgagatc 3'	70
			5' caatttcacacagcggatgcttct tttgctctgact 3'	71
Prothrombin	26731 gaattatttt tgtgtttcta aaactatggt <u>tccaataaa agtgactctc</u> 26781 <u>agc[g&gt;a]agcctc aatgctccca</u> <u>gtgctattca tgggcagctc tctgggctca</u>	100-101	5' tcagtcacgacgttggatgcca taaaagtgactctcagc 3'	72
			5' cggataacaatttcggatgcact gggagcattgagc 3'	73
Factor V (Arg506Gln)	taataggact acttctaadc tgtaagagca <u>gatccctgga caggc[g&gt;a]agga</u>  <u>atacaggtat ttgtccttg aagtaacctt tcag</u>	102-103	5' tcagtcacgacgttggatgagca gatccctggacagc 3'	74
			5' cggataacaatttcggatggaca aaataacctgtattcc 3'	75
Cholesterol ester transfer protein (CETP) (I405V)	1261 <u>ctcaccatgg gcatttgatt gcagagcage</u> <u>tccgagctcc[g&gt;a] tccagagctt</u> 1311 <u>cctgcagtica atgatacccg ctgtgggcat</u> ccctgaggtc atgtctcgta	104-105	5' tcagtcacgacgttggatgcaga gcagctccgagtc 3'	76
			5' <u>cagcgggtgatcattggatgcagg</u> <u>aagctctgg 3'</u>	77
Factor VII (R353Q)	1221 agcaaggact cctgcaaggg ggacagtggg <u>ggccacacatg ccacccacta</u> 1271 <u>cc[a&gt;g]gggcacg tggtaacctga</u> <u>cgggcatcgt cagctggggc cagggtcgcg</u>	106-107	5' tcagtcacgacgttggatgccca catgccacccactac 3'	78
			5' cggataacaatttcggatgcccq tcaggatccacg 3'	79
Factor XIII (V34L)	111 caataactct aatgcagcgg aagatgacct <u>gcccacagtg gacgttcagg</u> 161 <u>gc[g&gt;t]tgggtgcc ccggggcgtc</u> <u>aacctgcaag gtatgagcat accccccttc</u>	108-109	5' tcagtcacgacgttggatgccca cagtggagcttcag 3'	80
			5' <u>gctcataccttgaggatgacg</u> 3'	81
HLA-H exon 2 (His63Asp)	361 ttgaagcttt gggctacgtg <u>gatgaccagc</u> <u>gttcgtgtt ctatgat[c&gt;g]at</u> 411 <u>gagagtcgcc gtgtggagcc ccgaactcca</u> <u>tgggtttcca gtagaatttc</u>	110-111	5' tcagtcacgacgttggatgacca gctgttcgtgttc 3'	82
			5' tacatggagttcggggatgcaca cggcgactctc 3'	83
HLA-H exon 4 (Cys282Tyr)	1021 ggataacctt ggctgtaccc <u>cctggggaag</u> <u>agcagagata tacgt[g&gt;a]lccag</u> 1071 <u>gtggagcacc caggcctgga tcagcccttc</u> attgtgatct gggagccctc	112-113	5' tcagtcacgacgttggatgggga agagcagagatatacgt 3'	84
			5' <u>gaggggctgatccaggatgggt</u> <u>gtccac 3'</u>	85

5

Gene	Sequence	Seq. ID No.	Primers	Seq. ID No.
Methylenetetrahydrofolate reductase (MTHFR) (Ala222Val)	761 <u>tgaagcactt gaagga gaag gtgtctgcgg</u> gag(c>t) <u>cgattt catcatcacg</u>	114-115	5' tcagtcacgacgttggat <u>gggga agagcagagatatacgt</u> 3'	86
	811 <u>cagcttttct ttgaggctga cacattcttc</u>		5' <u>gaggggctgatccaggatgggt gctccac</u> 3'	87
P53 Exon4 Codon 72 (Arg72Pro)	12101 <u>tccagatgaa gctcccagaa</u> <u>tgccagaggc tgctccc(g&gt;c)c</u> gtggccctg	116-117	5' <u>gatgaagctcccaggatgccag aggc</u> 3'	88
	12151 <u>caccagcagc tcctacaccg</u> <u>gcggccctg</u>		5' <u>gccgccggtgtaggatgctgctg gtgc</u> 3'	89

Table 6  
The mass of Center Fragments for Ten Different SNP Typing by  
IIS Assay

Gene	LPL <sup>(Ala291Ser)</sup>		Prothrombin		FV <sup>(Arg506Gln)</sup>		CETP <sup>(405Val)</sup>		FVII <sup>(353G)</sup>		FXIII <sup>(34T)</sup>	
Genotype	A	G	G	A	G	A	G	A	G	A	G	T
+ strand mass (Da)	6213	6229	5845	5829	5677	5661	3388	3372	6128	6112	5058	5033
- strand mass (Da)	6129	6114	5949	5964	5472	5487	3437	3452	6174	6189	4916	4940

Gene	H1ah2		H1ah4		MTHR <sup>(Ala222Val)</sup>		P53exon4 <sup>(Ala72Pro)</sup>	
Genotype	C	G	G	A	C	T	G	C
+ strand mass (Da)	5889	5929	4392	4376	4400	4415	4586	4546
- strand mass (Da)	5836	5796	4319	4334	4368	4352	4724	4764

### EXAMPLE 10

#### Exemplary use of parental medical history parameter for stratification of healthy database

A healthy database can be used to associate a disease state with a specific allele (SNP) that has been found to show a strong association between age and the allele, in particular the homozygous genotype. The method involves using the same healthy database used to identify the age dependent association, however stratification is by information given by the donors about common disorders from which their parents suffered (the donor's familial history of disease). There are three possible answers a donor could give about the health status of their parents: neither were affected, one was affected or both were affected. Only donors above a certain minimum age, depending on the disease, are utilized, as the donors parents must be old enough to to have exhibited clinical disease phenotypes. The genotype frequency in each of these groups is determined and compared with each other. If there is an association of the marker in the donor to a disease the frequency of the heterozygous genotype will be increased. The frequency of the homozygous genotype should not increase, as it should be significantly underrepresented in the healthy population.

**EXAMPLE 11****Method and Device for Identifying a Biological Sample****Description**

In accordance with the present invention, a method and device for  
5 identifying a biological sample is provided. Referring now to FIG. 24, an  
apparatus 10 for identifying a biological sample is disclosed. The apparatus 10  
for identifying a biological sample generally comprises a mass spectrometer 15  
communicating with a computing device 20. In a preferred embodiment, the  
mass spectrometer may be a MALDI-TOF mass spectrometer manufactured by  
10 Bruker-Franzen Analytik GmbH; however, it will be appreciated that other mass  
spectrometers can be substituted. The computing device 20 is preferably a  
general purpose computing device. However, it will be appreciated that the  
computing device could be alternatively configured, for example, it may be  
integrated with the mass spectrometer or could be part of a computer in a larger  
15 network system.

The apparatus 10 for identifying a biological sample may operate as an  
automated identification system having a robot 25 with a robotic arm 27  
configured to deliver a sample plate 29 into a receiving area 31 of the mass  
spectrometer 15. In such a manner, the sample to be identified may be placed  
20 on the plate 29 and automatically received into the mass spectrometer 15. The  
biological sample is then processed in the mass spectrometer to generate data  
indicative of the mass of DNA fragments in the biological sample. This data may  
be sent directly to computing device 20, or may have some preprocessing or  
filtering performed within the mass spectrometer. In a preferred embodiment,  
25 the mass spectrometer 15 transmits unprocessed and unfiltered mass  
spectrometry data to the computing device 20. However, it will be appreciated  
that the analysis in the computing device may be adjusted to accommodate  
preprocessing or filtering performed within the mass spectrometer.

Referring now to FIG. 25, a general method 35 for identifying a biological  
30 sample is shown. In method 35, data is received into a computing device from a  
test instrument in block 40. Preferably the data is received in a raw,  
unprocessed and unfiltered form, but alternatively may have some form of

filtering or processing applied. The test instrument of a preferred embodiment is a mass spectrometer as described above. However, it will be appreciated that other test instruments could be substituted for the mass spectrometer.

The data generated by the test instrument, and in particular the mass spectrometer, includes information indicative of the identification of the biological sample. More specifically, the data is indicative of the DNA composition of the biological sample. Typically, mass spectrometry data gathered from DNA samples obtained from DNA amplification techniques are noisier than, for example, those from typical protein samples. This is due in part because protein samples are more readily prepared in more abundance, and protein samples are more easily ionizable as compared to DNA samples. Accordingly, conventional mass spectrometer data analysis techniques are generally ineffective for DNA analysis of a biological sample. To improve the analysis capability so that DNA composition data can be more readily discerned, a preferred embodiment uses wavelet technology for analyzing the DNA mass spectrometry data. Wavelets are an analytical tool for signal processing, numerical analysis, and mathematical modeling. Wavelet technology provides a basic expansion function which is applied to a data set. Using wavelet decomposition, the data set can be simultaneously analyzed in the time and frequency domains. Wavelet transformation is the technique of choice in the analysis of data that exhibit complicated time (mass) and frequency domain information, such as MALDI-TOF DNA data. Wavelet transforms as described herein have superior denoising properties as compared to conventional Fourier analysis techniques. Wavelet transformation has proven to be particularly effective in interpreting the inherently noisy MALDI-TOF spectra of DNA samples. In using wavelets, a "small wave" or "scaling function" is used to transform a data set into stages, with each stage representing a frequency component in the data set. Using wavelet transformation, mass spectrometry data can be processed, filtered, and analyzed with sufficient discrimination to be useful for identification of the DNA composition for a biological sample.

Referring again to FIG. 25, the data received in block 40 is denoised in block 45. The denoised data then has a baseline correction applied in block 50.

A baseline correction is generally necessary as data coming from the test instrument, in particular a mass spectrometer instrument, has data arranged in a generally exponentially decaying manner. This generally exponential decaying arrangement is not due to the composition of the biological sample, but is a  
5 result of the physical properties and characteristics of the test instrument, and other chemicals involved in DNA sample preparation. Accordingly, baseline correction substantially corrects the data to remove a component of the data attributable to the test system, and sample preparation characteristics.

After denoising in block 45 and the baseline correction in block 50, a  
10 signal remains which is generally indicative of the composition of the biological sample. However, due to the extraordinary discrimination required for analyzing the DNA composition of the biological sample, the composition is not readily apparent from the denoised and corrected signal. For example, although the signal may include peak areas, it is not yet clear whether these "putative" peaks  
15 actually represent a DNA composition, or whether the putative peaks are result of a systemic or chemical aberration. Further, any call of the composition of the biological sample would have a probability of error which would be unacceptable for clinical or therapeutic purposes. In such critical situations, there needs to be a high degree of certainty that any call or identification of the sample is  
20 accurate. Therefore, additional data processing and interpretation is necessary before the sample can be accurately and confidently identified.

Since the quantity of data resulting from each mass spectrometry test is typically thousands of data points, and an automated system may be set to perform hundreds or even thousands of tests per hour, the quantity of mass  
25 spectrometry data generated is enormous. To facilitate efficient transmission and storage of the mass spectrometry data, block 55 shows that the denoised and baseline corrected data is compressed.

In a preferred embodiment, the biological sample is selected and processed to have only a limited range of possible compositions. Accordingly, it  
30 is therefore known where peaks indicating composition should be located, if present. Taking advantage of knowing the location of these expected peaks, in block 60 the method 35 matches putative peaks in the processed signal to the

location of the expected peaks. In such a manner, the probability of each putative peak in the data being an actual peak indicative of the composition of the biological sample can be determined. Once the probability of each peak is determined in block 60, then in block 65 the method 35 statistically determines the composition of the biological sample, and determines if confidence is high enough to calling a genotype.

Referring again to block 40, data is received from the test instrument, which is preferably a mass spectrometer. In a specific illustration, FIG. 26 shows an example of data from a mass spectrometer. The mass spectrometer data 70 generally comprises data points distributed along an x-axis 71 and a y-axis 72. The x-axis 71 represents the mass of particles detected, while the y-axis 72 represents a numerical concentration of the particles. As can be seen in FIG. 26, the mass spectrometry data 70 is generally exponentially decaying with data at the left end of the x-axis 73 generally decaying in an exponential manner toward data at the heavier end 74 of the x-axis 71. However, the general exponential presentation of the data is not indicative of the composition of the biological sample, but is more reflective of systematic error and characteristics. Further, as described above and illustrated in FIG. 26, considerable noise exists in the mass spectrometry DNA data 70.

Referring again to block 45, where the raw data received in block 40 is denoised, the denoising process will be described in more detail. As illustrated in FIG. 25, the denoising process generally entails 1) performing a wavelet transformation on the raw data to decompose the raw data into wavelet stage coefficients; 2) generating a noise profile from the highest stage of wavelet coefficients; and 3) applying a scaled noise profile to other stages in the wavelet transformation. Each step of the denoising process is further described below.

Referring now to FIG. 27, the wavelet transformation of the raw mass spectrometry data is generally diagramed. Using wavelet transformation techniques, the mass spectrometry data 70 is sequentially transformed into stages. In each stage the data is represented in a high stage and a low stage, with the low stage acting as the input to the next sequential stage. For example, the mass spectrometry data 70 is transformed into stage 0 high data

82 and stage 0 low data 83. The stage 0 low data 83 is then used as an input to the next level transformation to generate stage 1 high data 84 and stage 1 low data 85. In a similar manner, the stage 1 low data 85 is used as an input to be transformed into stage 2 high data 86 and stage 2 low data 87. The transformation is continued until no more useful information can be derived by further wavelet transformation. For example, in the preferred embodiment a 24-point wavelet is used. More particularly a wavelet commonly referred to as the Daubechies 24 is used to decompose the raw data. However, it will be appreciated that other wavelets can be used for the wavelet transformation.

10 Since each stage in a wavelet transformation has one-half the data points of the previous stage, the wavelet transformation can be continued until the stage n low data 89 has around 50 points. Accordingly, the stage n high 88 would contain about 100 data points. Since the preferred wavelet is 24 points long, little data or information can be derived by continuing the wavelet transformation

15 on a data set of around 50 points.

FIG. 28 shows an example of stage 0 high data 95. Since stage 0 high data 95 is generally indicative of the highest frequencies in the mass spectrometry data, stage 0 high data 95 will closely relate to the quantity of high frequency noise in the mass spectrometry data. In FIG. 29, an exponential fitting formula has been applied to the stage 0 high data 95 to generate a stage 0 noise profile 97. In particular, the exponential fitting formula is in the format  $A_0 + A_1 \text{EXP}(-A_2 m)$ . It will be appreciated that other exponential fitting formulas or other types of curve fits may be used.

20

Referring now to FIG. 30, noise profiles for the other high stages are determined. Since the later data points in each stage will likely be representative of the level of noise in each stage, only the later data points in each stage are used to generate a standard deviation figure that is representative of the noise content in that particular stage. More particularly, in generating the noise profile for each remaining stage, only the last five percent of the data points in each stage are analyzed to determine a standard deviation number. It will be appreciated that other numbers of points, or alternative methods could be used to generate such a standard deviation figure.

25

30

The standard deviation number for each stage is used with the stage 0 noise profile (the exponential curve) 97 to generate a scaled noise profile for each stage. For example, FIG. 30 shows that stage 1 high data 98 has stage 1 high data 103 with the last five percent of the data points represented by area 5 99. The points in area 99 are evaluated to determine a standard deviation number indicative of the noise content in stage 1 high data 103. The standard deviation number is then used with the stage 0 noise profile 97 to generate a stage 1 noise profile.

In a similar manner, stage 2 high 100 has stage 2 high data 104 with the 10 last five percent of points represented by area 101. The data points in area 101 are then used to calculate a standard deviation number which is then used to scale the stage 0 noise profile 97 to generate a noise profile for stage 2 data. This same process is continued for each of the stage high data as shown by the stage n high 105. For stage n high 105, stage n high data 108 has the last five 15 percent of data points indicated in area 106. The data points in area 106 are used to determine a standard deviation number for stage n. The stage n standard deviation number is then used with the stage 0 noise profile 97 to generate a noise profile for stage n. Accordingly, each of the high data stages has a noise profile.

20 FIG. 31 shows how the noise profile is applied to the data in each stage. Generally, the noise profile is used to generate a threshold which is applied to the data in each stage. Since the noise profile is already scaled to adjust for the noise content of each stage, calculating a threshold permits further adjustment to tune the quantity of noise removed. Wavelet coefficients below the threshold 25 are ignored while those above the threshold are retained. Accordingly, the remaining data has a substantial portion of the noise content removed.

Due to the characteristics of wavelet transformation, the lower stages, such as stage 0 and 1, will have more noise content than the later stages such as stage 2 or stage n. Indeed, stage n low data is likely to have little noise at 30 all. Therefore, in a preferred embodiment the noise profiles are applied more aggressively in the lower stages and less aggressively in the later stages. For example, FIG. 31 shows that stage 0 high threshold is determined by multiplying

the stage 0 noise profile by a factor of four. In such a manner, significant numbers of data points in stage 0 high data 95 will be below the threshold and therefore eliminated. Stage 1 high threshold 112 is set at two times the noise profile for the stage 1 high data, and stage 2 high threshold 114 is set equal to the noise profile for stage 2 high. Following this geometric progression, stage n high threshold 116 is therefore determined by scaling the noise profile for each respective stage n high by a factor equal to  $(1/2^{n-2})$ . It will be appreciated that other factors may be applied to scale the noise profile for each stage. For example, the noise profile may be scaled more or less aggressively to accommodate specific systemic characteristics or sample compositions. As indicated above, stage n low data does not have a noise profile applied as stage n low data 118 is assumed to have little or no noise content. After the scaled noise profiles have been applied to each high data stage, the mass spectrometry data 70 has been denoised and is ready for further processing. A wavelet transformation of the denoised signal results in the sparse data set 120 as shown in FIG. 31.

Referring again to FIG. 25, the mass spectrometry data received in block 40 has been denoised in block 45 and is now passed to block 50 for baseline correction. Before performing baseline correction, the artifacts introduced by the wavelet transformation procedure are preferably removed. Wavelet transformation results vary slightly depending upon which point of the wavelet is used as a starting point. For example, the preferred embodiment uses the 24-point Daubechies-24 wavelet. By starting the transformation at the 0 point of the wavelet, a slightly different result will be obtained than if starting at points 1 or 2 of the wavelet. Therefore, the denoised data is transformed using every available possible starting point, with the results averaged to determine a final denoised and shifted signal. For example, FIG. 33 shows that the wavelet coefficient is applied 24 different times and then the results averaged to generate the final data set. It will be appreciated that other techniques may be used to accommodate the slight error introduced due to wavelet shifting.

The formula 125 is generally indicated in FIG. 33. Once the signal has been denoised and shifted, a denoised and shifted signal 130 is generated as

shown in FIG. 58. FIG. 34 shows an example of the wavelet coefficient 135 data set from the denoised and shifted signal 130.

FIG. 36 shows that putative peak areas 145, 147, and 149 are located in the denoised and shifted signal 150. The putative peak areas are systematically  
5 identified by taking a moving average along the signal 150 and identifying sections of the signal 150 which exceed a threshold related to the moving average. It will be appreciated that other methods can be used to identify putative peak areas in the signal 150.

Putative peak areas 145, 147 and 149 are removed from the signal 150  
10 to create a peak-free signal 155 as shown in FIG. 37. The peak-free signal 155 is further analyzed to identify remaining minimum values 157, and the remaining minimum values 157 are connected to generate the peak-free signal 155.

FIG. 38 shows a process of using the peak-free signal 155 to generate a baseline 170 as shown in FIG. 39. As shown in block 162, a wavelet  
15 transformation is performed on the peak-free signal 155. All the stages from the wavelet transformation are eliminated in block 164 except for the n low stage. The n low stage will generally indicate the lowest frequency component of the peak-free signal 155 and therefore will generally indicate the system exponential characteristics. Block 166 shows that a signal is reconstructed from the n low  
20 coefficients and the baseline signal 170 is generated in block 168.

FIG. 39 shows a denoised and shifted data signal 172 positioned adjacent a correction baseline 170. The baseline correction 170 is subtracted from the denoised and shifted signal 172 to generate a signal 175 having a baseline correction applied as shown in FIG. 40. Although such a denoised, shifted, and  
25 corrected signal is sufficient for most identification purposes, the putative peaks in signal 175 are not identifiable with sufficient accuracy or confidence to call the DNA composition of a biological sample.

Referring again to FIG. 25, the data from the baseline correction 50 is now compressed in block 55, the compression technique used in a preferred  
30 embodiment is detailed in FIG. 41. In FIG. 41 the data in the baseline corrected data is presented in an array format 182 with x-axis points 183 having an associated data value 184. The x-axis is indexed by the non-zero wavelet

coefficients, and the associated value is the value of the wavelet coefficient. In the illustrated data example in table 182, the maximum value 184 is indicated to be 1000. Although a particularly advantageous compression technique for mass spectrometry data is shown, it will be appreciated that other compression techniques can be used. Although not preferred, the data may also be stored without compression.

In compressing the data according to a preferred embodiment, an intermediate format 186 is generated. The intermediate format 186 generally comprises a real number having a whole number portion 188 and a decimal portion 190. The whole number portion is the x-axis point 183 while the decimal portion is the value data 184 divided by the maximum data value. For example, in the data 182 a data value "25" is indicated at x-axis point "100". The intermediate value for this data point would be "100.025".

From the intermediate compressed data 186 the final compressed data 195 is generated. The first point of the intermediate data file becomes the starting point for the compressed data. Thereafter each data point in the compressed data 195 is calculated as follows: the whole number portion (left of the decimal) is replaced by the difference between the current and the last whole number. The remainder (right of the decimal) remains intact. For example, the starting point of the compressed data 195 is shown to be the same as the intermediate data point which is "100.025". The comparison between the first intermediate data point "100.025" and the second intermediate data point "150.220" is "50.220". Therefore, "50.220" becomes the second point of the compressed data 195. In a similar manner, the second intermediate point is "150.220" and the third intermediate data point is "500.0001". Therefore, the third compressed data becomes "350.000". The calculation for determining compressed data points is continued until the entire array of data points is converted to a single array of real numbers.

FIG. 42 generally describes the method of compressing mass spectrometry data, showing that the data file in block 201 is presented as an array of coefficients in block 202. The data starting point and maximum is determined as shown in block 203, and the intermediate real numbers are

calculated in block 204 as described above. With the intermediate data points generated, the compressed data is generated in block 205. The described compression method is highly advantageous and efficient for compressing data sets such as a processed data set from a mass spectrometry instrument. The  
5 method is particularly useful for data, such as mass spectrometry data, that uses large numbers and has been processed to have occasional lengthy gaps in x-axis data. Accordingly, an x-y data array for processed mass spectrometry data may be stored with an effective compression rate of 10x or more. Although the compression technique is applied to mass spectrometry data, it will be  
10 appreciated that the method may also advantageously be applied to other data sets.

Referring again to FIG. 25, peak heights are now determined in block 60. The first step in determining peak height is illustrated in FIG. 43 where the signal 210 is shifted left or right to correspond with the position of expected peaks.  
15 As the set of possible compositions in the biological sample is known before the mass spectrometry data is generated, the possible positioning of expected peaks is already known. These possible peaks are referred to as expected peaks, such as expected peaks 212, 214, and 216. Due to calibration or other errors in the test instrument data, the entire signal may be shifted left or right from its actual  
20 position, therefore, putative peaks located in the signal, such as putative peaks 218, 222, and 224 may be compared to the expected peaks 212, 214, and 216, respectively. The entire signal is then shifted such that the putative peaks align more closely with the expected peaks.

Once the putative peaks have been shifted to match expected peaks, the  
25 strongest putative peak is identified in FIG. 44. In a preferred embodiment, the strongest peak is calculated as a combination of analyzing the overall peak height and area beneath the peak. For example, a moderately high but wide peak would be stronger than a very high peak that is extremely narrow. With the strongest putative peak identified, such as putative peak 225, a Gaussian  
30 228 curve is fit to the peak 225. Once the Gaussian is fit, the width (W) of the Gaussian is determined and will be used as the peak width for future calculations.

As generally addressed above, the denoised, shifted, and baseline-corrected signal is not sufficiently processed for confidently calling the DNA composition of the biological sample. For example, although the baseline has generally been removed, there are still residual baseline effects present. These  
5 residual baseline effects are therefore removed to increase the accuracy and confidence in making identifications.

To remove the residual baseline effects, FIG. 45 shows that the putative peaks 218, 222, and 224 are removed from the baseline corrected signal. The peaks are removed by identifying a center line 230, 232, and 234 of the  
10 putative peaks 218, 222, and 224, respectively and removing an area to the left and to the right of the identified center line. For each putative peak, an area equal to twice the width (W) of the Gaussian is removed from the left of the center line, while an area equivalent to 50 daltons is removed from the right of the center line. It has been found that the area representing 50 daltons is  
15 adequate to sufficiently remove the effect of salt adducts which may be associated with an actual peak. Such adducts appear to the right of an actual peak and are a natural effect from the chemistry involved in acquiring a mass spectrum. Although a 50 Dalton buffer has been selected, it will be appreciated that other ranges or methods can be used to reduce or eliminate adduct effects.

20 The peaks are removed and remaining minima 247 located as shown in FIG. 46 with the minima 247 connected to create signal 245. A quartic polynomial is applied to signal 245 to generate a residual baseline 250 as shown in FIG. 47. The residual baseline 250 is subtracted from the signal 225 to generate the final signal 255 as indicated in FIG. 48. Although the residual  
25 baseline is the result of a quartic fit to signal 245, it will be appreciated that other techniques can be used to smooth or fit the residual baseline.

To determine peak height, as shown in FIG. 49, a Gaussian such as Gaussian 266, 268, and 270 is fit to each of the peaks, such as peaks 260, 262, and 264, respectively. Accordingly, the height of the Gaussian is  
30 determined as height 272, 274, and 276. Once the height of each Gaussian peak is determined, then the method of identifying a biological compound 35 can move into the genotyping phase 65 as shown in FIG. 25.

An indication of the confidence that each putative peak is an actual peak can be discerned by calculating a signal-to-noise ratio for each putative peak. Accordingly, putative peaks with a strong signal-to-noise ratio are generally more likely to be an actual peak than a putative peak with a lower signal-to-noise ratio. As described above and shown in FIG. 50, the height of each peak, such as height 272, 274, and 276, is determined for each peak, with the height being an indicator of signal strength for each peak. The noise profile, such as noise profile 97, is extrapolated into noise profile 280 across the identified peaks. At the center line of each of the peaks, a noise value is determined, such as noise value 282, 283, and 284. With a signal values and a noise values generated, signal-to-noise ratios can be calculated for each peak. For example, the signal-to-noise ratio for the first peak in FIG. 50 would be calculated as signal value 272 divided by noise value 282, and in a similar manner the signal-to-noise ratio of the middle peak in FIG. 50 would be determined as signal 274 divided by noise value 283.

Although the signal-to-noise ratio is generally a useful indicator of the presence of an actual peak, further processing has been found to increase the confidence by which a sample can be identified. For example, the signal-to-noise ratio for each peak in the preferred embodiment is preferably adjusted by the goodness of fit between a Gaussian and each putative peak. It is a characteristic of a mass spectrometer that sample material is detected in a manner that generally complies with a normal distribution. Accordingly, greater confidence will be associated with a putative signal having a Gaussian shape than a signal that has a less normal distribution. The error resulting from having a non-Gaussian shape can be referred to as a "residual error".

Referring to FIG. 51, a residual error is calculated by taking a root mean square calculation between the Gaussian 293 and the putative peak 290 in the data signal. The calculation is performed on data within one width on either side of a center line of the Gaussian. The residual error is calculated as:  $\sqrt{(G-R)^2/N}$  where G is the Gaussian signal value, R is the putative peak value, and N is the number of points from -W to +W. The calculated residual error is used to generate an adjusted signal-to-noise ratio, as described below.

An adjusted signal noise ratio is calculated for each putative peak using the formula  $(S/N) * \text{EXP}^{-1 * R}$ , where S/N is the signal-to-noise ratio, and R is the residual error determined above. Although the preferred embodiment calculates an adjusted signal-to-noise ratio using a residual error for each peak, it  
5 will be appreciated that other techniques can be used to account for the goodness of fit between the Gaussian and the actual signal.

Referring now to FIG. 52, a probability is determined that a putative peak is an actual peak. In making the determination of peak probability, a probability profile 300 is generated where the adjusted signal-to-noise ratio is the x-axis and  
10 the probability is the y-axis. Probability is necessarily in the range between a 0% probability and a 100% probability, which is indicated as 1. Generally, the higher the adjusted signal-to-noise ratio, the greater the confidence that a putative peak is an actual peak.

At some target value for the adjusted signal-to-noise, it has been found  
15 that the probability is 100% that the putative peak is an actual peak and can confidently be used to identify the DNA composition of a biological sample. However, the target value of adjusted signal-to-noise ratio where the probability is assumed to be 100% is a variable parameter which is to be set according to application specific criteria. For example, the target signal-to-noise ratio will be  
20 adjusted depending upon trial experience, sample characteristics, and the acceptable error tolerance in the overall system. More specifically, for situations requiring a conservative approach where error cannot be tolerated, the target adjusted signal-to-noise ratio can be set to, for example, 10 and higher. Accordingly, 100% probability will not be assigned to a peak unless the adjusted  
25 signal-to-noise ratio is 10 or over.

In other situations, a more aggressive approach may be taken as sample data is more pronounced or the risk of error may be reduced. In such a situation, the system may be set to assume a 100% probability with a 5 or greater target signal-to-noise ratio. Of course, an intermediate signal-to-noise  
30 ratio target figure can be selected, such as 7, when a moderate risk of error can be assumed. Once the target adjusted signal-to-noise ratio is set for the method,

then for any adjusted signal-to-noise ratio a probability can be determined that a putative peak is an actual peak.

Due to the chemistry involved in performing an identification test, especially a mass spectrometry test of a sample prepared by DNA amplifications, the allelic ratio between the signal strength of the highest peak and the signal strength of the second (or third and so on) highest peak should fall within an expected ratio. If the allelic ratio falls outside of normal guidelines, the preferred embodiment imposes an allelic ratio penalty to the probability. For example, FIG. 53 shows an allelic penalty 315 which has an x-axis 317 that is the ratio between the signal strength of the second highest peak divided by signal strength of the highest peak. The y-axis 319 assigns a penalty between 0 and 1 depending on the determined allelic ratio. In the preferred embodiment, it is assumed that allelic ratios over 30% are within the expected range and therefore no penalty is applied. Between a ratio of 10% and 30%, the penalty is linearly increased until at allelic ratios below 10% it is assumed the second-highest peak is not real. For allelic ratios between 10% and 30%, the allelic penalty chart 315 is used to determine a penalty 319, which is multiplied by the peak probability determined in FIG. 52 to determine a final peak probability. Although the preferred embodiment incorporates an allelic ratio penalty to account for a possible chemistry error, it will be appreciated that other techniques may be used. Similar treatment will be applied to the other peaks.

With the peak probability of each peak determined, the statistical probability for various composition components may be determined. As an example, in order to determine the probability of each of three possible combinations of two peaks, -- peak G, peak C and combinations GG, CC and GC. FIG. 54 shows an example where a most probable peak 325 is determined to have a final peak probability of 90%. Peak 325 is positioned such that it represents a G component in the biological sample. Accordingly, it can be maintained that there is a 90% probability that G exists in the biological sample. Also in the example shown in FIG. 54, the second highest probability is peak 330 which has a peak probability of 20%. Peak 330 is at a position associated

with a C composition. Accordingly, it can be maintained that there is a 20% probability that C exists in the biological sample.

With the probability of G existing (90%) and the probability of C existing (20%) as a starting point, the probability of combinations of G and C existing  
5 can be calculated. For example, FIG. 54 indicates that the probability of GG existing 329 is calculated as 72%. This is calculated as the probability of GG is equal to the probability of G existing (90%) multiplied by the probability of C not existing (100% - 20%). So if the probability of G existing is 90% and the probability of C not existing is 80%, the probability of GG is 72%.

10 In a similar manner, the probability of CC existing is equivalent to the probability of C existing (20%) multiplied by the probability of G not existing (100% - 90%). As shown in FIG. 54, the probability of C existing is 20% while the probability of G not existing is 10%, so therefore the probability of CC is only 2%. Finally, the probability of GC existing is equal to the probability of G  
15 existing (90%) multiplied by the probability of C existing (20%). So if the probability of G existing is 90% and the probability of C existing is 20%, the probability of GC existing is 18%. In summary form, then, the probability of the composition of the biological sample is:

20 probability of GG: 72%;  
probability of GC: 18%; and  
probability of CC: 2%.

Once the probabilities of each of the possible combinations has been determined, FIG. 55 is used to decide whether or not sufficient confidence exists to call the genotype. FIG. 55 shows a call chart 335 which has an x-axis 337  
25 which is the ratio of the highest combination probability to the second highest combination probability. The y-axis 339 simply indicates whether the ratio is sufficiently high to justify calling the genotype. The value of the ratio may be indicated by M 340. The value of M is set depending upon trial data, sample composition, and the ability to accept error. For example, the value M may be  
30 set relatively high, such as to a value 4 so that the highest probability must be at least four times greater than the second highest probability before confidence is established to call a genotype. However, if a certain level of error may be

acceptable, the value of M may be set to a more aggressive value, such as to 3, so that the ratio between the highest and second highest probabilities needs to be only a ratio of 3 or higher. Of course, moderate value may be selected for M when a moderate risk can be accepted. Using the example of FIG. 54, where  
5 the probability of GG was 72% and the probability of GC was 18%, the ratio between 72% and 18% is 4.0, therefore, whether M is set to 3, 3.5, or 4, the system would call the genotype as GG. Although the preferred embodiment uses a ratio between the two highest peak probabilities to determine if a genotype confidently can be called, it will be appreciated that other methods  
10 may be substituted. It will also be appreciated that the above techniques may be used for calculating probabilities and choosing genotypes (or more general DNA patterns) containing of combinations of more than two peaks.

Referring now to FIG. 56, a flow chart is shown generally defining the process of statistically calling genotype described above. In FIG. 56 block 402  
15 shows that the height of each peak is determined and that in block 404 a noise profile is extrapolated for each peak. The signal is determined from the height of each peak in block 406 and the noise for each peak is determined using the noise profile in block 408. In block 410, the signal-to-noise ratio is calculated for each peak. To account for a non-Gaussian peak shape, a residual error is  
20 determined in block 412 and an adjusted signal-to-noise ratio is calculated in block 414. Block 416 shows that a probability profile is developed, with the probability of each peak existing found in block 418. An allelic penalty may be applied in block 420, with the allelic penalty applied to the adjusted peak probability in block 422. The probability of each combination of components is  
25 calculated in block 424 with the ratio between the two highest probabilities being determined in block 426. If the ratio of probabilities exceeds a threshold value then the genotype is called in block 428.

In another embodiment of the invention, the computing device 20 (Fig.  
30 24) supports "standardless" genotyping by identifying data peaks that contain putative SNPs. Standardless genotyping is used, for example, where insufficient information is known about the samples to determine a distribution of expected

peak locations, against which an allelic penalty as described above can be reliably calculated. This permits the computing device to be used for identification of peaks that contain putative SNPs from data generated by any assay that fragments a targeted DNA molecule. For such standardless

- 5 genotyping, peaks that are associated with an area under the data curve that deviates significantly from the typical area of other peaks in the data spectrum are identified and their corresponding mass (location along the x-axis) is determined.

- More particularly, peaks that deviate significantly from the average area  
10 of other peaks in the data are identified, and the expected allelic ratio between data peaks is defined in terms of the ratio of the area under the data peaks. Theoretically, where each genetic loci has the same molar concentration of analyte, the area under each corresponding peak should be the same, thus producing a 1.0 ratio of the peak area between any two peaks. In accordance  
15 with the invention, peaks having a smaller ratio relative to the other peaks in the data will not be recognized as peaks. More particularly, peaks having an area ratio smaller than 30% relative to a nominal value for peak area will be assigned an allelic penalty. The mass of the remaining peaks (their location along the x-axis of the data) will be determined based on oligonucleotide standards.

- 20 Fig. 57 shows a flow diagram representation of the processing by the computing device 20 (Fig. 24) when performing standardless genotyping. In the first operation, represented by the flow diagram box numbered 502, the computing device receives data from the mass spectrometer. Next, the height of each putative peak in the data sample is determined, as indicated by the block  
25 504. After the height of each peak in the mass spectrometer data is determined, a de-noise process 505 is performed, beginning with an extrapolation of the noise profile (block 506), followed by finding the noise of each peak (block 508) and calculating the signal to noise ratio for each data sample (block 510). Each of these operations may be performed in accordance  
30 with the description above for denoise operations 45 of Fig. 25. Other suitable denoise operations will occur to those skilled in the art.

The next operation is to find the residual error associated with each data point. This is represented by the block 512 in Figure 57. The next step, block 514, involves calculating an adjusted signal to noise ratio for each identified peak. A probability profile is developed next (block 516), followed by a  
5 determination of the peak probabilities at block 518. In the preferred embodiment, the denoise operations of Fig. 57, comprising block 502 to block 518, comprise the corresponding operations described above in conjunction with Fig. 56 for block 402 through block 418, respectively.

The next action for the standardless genotype processing is to determine  
10 an allelic penalty for each peak, indicated by the block 524. As noted above, the standardless genotype processing of Fig. 57 determines an allelic penalty by comparing area under the peaks. Therefore, rather than compare signal strength ratios to determine an allelic penalty, such as described above for Fig. 53, the standardless processing determines the area under each of the identified peaks  
15 and compares the ratio of those areas. Determining the area under each peak may be computed using conventional numerical analysis techniques for calculating the area under a curve for experimental data.

Thus, the allelic penalty is assigned in accordance with Fig. 58, which shows that no penalty is assigned to peaks having a peak area relative to an  
20 expected average area value that is greater than 0.30 (30%). The allelic penalty is applied to the peak probability value, which may be determined according to the process such as described in Fig. 52. It should be apparent from Fig. 58 that the allelic penalty imposed for peaks below a ratio of 30% is that such peaks will be removed from further measurement and processing. Other penalty  
25 schemes, however, may be imposed in accordance with knowledge about the data being processed, as determined by those skilled in the art.

After the allelic penalty has been determined and applied, the standardless genotype processing compares the location of the remaining putative peaks to oligonucleotide standards to determine corresponding masses  
30 in the processing for block 524. For standardless genotype data, the processing of the block 524 is performed to determine mass and genotype, rather than performing the operations corresponding to block 424, 426, and 428 of Fig. 33.

Techniques for performing such comparisons and determining mass will be known to those skilled in the art.

In another embodiment, the computing device 20 (Fig. 24) permits the detection and determination of the mass (location along the x-axis of the data) of the sense and antisense strand of fragments generated in the assay. If desired, the computing device may also detect and determine the quantity (area under each peak) of the respective sense and antisense strands, using a similar technique to that described above for standardless genotype processing. The data generated for each type of strand may then be combined to achieve a data redundancy and to thereby increase the confidence level of the determined genotype. This technique obviates primer peaks that are often observed in data from other diagnostic methods, thereby permitting a higher level of multiplexing. In addition, when quantitation is used in pooling experiments, the ratio of the measured peak areas is more reliably calculated than the peak identifying technique, due to data redundancy.

Fig. 23 is a flow diagram that illustrates the processing implemented by the computing device 20 to perform sense and antisense processing. In the first operation, represented by the flow diagram box numbered 602, the computing device receives data from the mass spectrometer. This data will include data for the sense strand and antisense strand of assay fragments. Next, the height of each putative peak in the data sample is determined, as indicated by the block 604. After the height of each peak in the mass spectrometer data is determined, a de-noise process 605 is performed, beginning with an operation that extrapolates the noise profile (block 606), followed by finding the noise of each peak (block 608) and calculating the signal to noise ratio for each data sample (block 610). Each of these operations may be performed in accordance with the description above for the denoise operations 45 of Fig. 25. Other suitable denoise operations will occur to those skilled in the art. The next operation is to find the residual error associated with each data point. This is represented by the block 612 in Figure 36.

After the residual error for the data of the sense strand and antisense strand has been performed, processing to identify the genotypes will be

performed for the sense strand and also for the antisense strand. Therefore, Fig. 23 shows that processing includes sense strand processing (block 630) and antisense strand processing (block 640). Each block 630, 640 includes processing that corresponds to adjusting the signal to noise ratio, developing a probability profile, determining an allelic penalty, adjusting the peak probability by the allelic penalty, calculating genotype probabilities, and testing genotype probability ratios, such as described above in conjunction with blocks 414 through 426 of Fig. 56. The processing of each block 630, 640 may, if desired, include standardless processing operations such as described above in conjunction with Fig. 57. The standardless processing may be included in place of or in addition to the processing operations of Fig. 56.

After the genotype probability processing is completed, the data from the sense strand and antisense strand processing is combined and compared to expected database values to obtain the benefits of data redundancy as between the sense strand and antisense strand. Those skilled in the art will understand techniques to take advantage of known data redundancies between a sense strand and antisense strand of assay fragments. This processing is represented by the block 650. After the data from the two strands is combined for processing, the genotype processing is performed (block 660) and the genotype is identified.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

**WHAT IS CLAIMED IS:**

1. A subcollection of samples from a target population, comprising:  
a plurality of samples, wherein the samples are selected from the group  
consisting of blood, tissue, body fluid, cell, seed, microbe, pathogen and  
5 reproductive tissue samples; and  
a symbology on the containers containing the samples, wherein the  
symbology is representative of the source and/or history of each sample,  
wherein:  
the target population is a healthy population that has not been selected  
10 for any disease state;  
the collection comprises samples from the healthy population; and  
the subcollection is obtained by sorting the collection according to  
specified parameters.
2. The subcollection of claim 1, wherein the parameters are selected  
15 from the group consisting of ethnicity, age, gender, height, weight, alcohol  
intake, number of pregnancies, number of live births, vegetarians, type of  
physical activity, state of residence and/or length of residence in a particular  
state, educational level, age of parent at death, cause of parent death, former or  
current smoker, length of time as a smoker, frequency of smoking, occurrence  
20 of a disease in immediate family (parent, siblings, children), use of prescription  
drugs and/or reason therefor, length and/or number of hospital stays and  
exposure to environmental factors.
3. The subcollection of claim 1, wherein the symbology is a bar code.
4. A method of producing a database, comprising:  
25 identifying healthy members of a population;  
obtaining data comprising identifying information and obtaining historical  
information and data relating to the identified members of the population and  
their immediate family;  
entering the data into a database for each member of the population and  
30 associating the member and the data with an indexer.
5. The method of claim 4, further comprising:  
obtaining a body tissue or body fluid sample;

analyzing the body tissue or body fluid in the sample; and  
entering the results of the analysis for each member into the database  
and associating each result with the indexer representative of each member.

6. A database produced by the method of claim 4.

5 7. A database produced by the method of claim 5.

8. A database, comprising:

datapoints representative of a plurality of healthy organisms from  
whom biological samples are obtained,

10 wherein each datapoint is associated with data representative of  
the organism type and other identifying information.

9. The database of claim 8, wherein the datapoints are answers to  
questions regarding one or more of a parameters selected from the group  
consisting of ethnicity, age, gender, height, weight, alcohol intake, number of  
pregnancies, number of live births, vegetarians, type of physical activity, state of  
15 residence and/or length of residence in a particular state, educational level, age  
of parent at death, cause of parent death, former or current smoker, length of  
time as a smoker, frequency of smoking, occurrence of a disease in immediate  
family (parent, siblings, children), use of prescription drugs and/or reason  
therefor, length and/or number of hospital stays and exposure to environmental  
20 factors.

10. The database of claim 9, wherein the organisms are mammals and  
the samples are body fluids or tissues.

11. The database of claim 9, wherein the samples are selected from  
blood, blood fractions, cells and subcellular organelles.

25 12. The database of claim 8, further comprising,  
phenotypic data from an organism.

13. The database of claim 12, wherein the data includes one of physical  
characteristics, background data, medical data, and historical data.

30 14. The database of claim 8, further comprising,  
genotypic data from nucleic acid obtained from an organism.

15. The database of claim 14, wherein genotypic data includes, genetic markers, non-coding regions, microsatellites, RFLPs, VNTRs, historical data of the organism, medical history, and phenotypic information.

16. The database of claim 8 that is a relational database.

5 17. The database of claim 16, wherein the data are related to an indexer datapoint representative of each organism from whom data is obtained.

18. A method of identifying polymorphisms that are candidate genetic markers, comprising:

identifying a polymorphism; and

10 identifying any pathway or gene linked to the locus of the polymorphism, wherein

the polymorphisms are identified in samples associated with a target population that comprises healthy subjects.

19. The method of claim 18, wherein the polymorphism is identified by  
15 detecting the presence of target nucleic acids in a sample by a method, comprising the steps of:

a) hybridizing a first oligonucleotide to the target nucleic acid;

b) hybridizing a second oligonucleotide to an adjacent region of the target nucleic acid;

20 c) ligating the hybridized oligonucleotides; and

c) detecting hybridized first oligonucleotide by mass spectrometry as an indication of the presence of the target nucleic acid.

20. The method of claim 18, wherein the polymorphism is identified by detecting target nucleic acids in a sample by a method, comprising the steps of:

25 a) hybridizing a first oligonucleotide to the target nucleic acid and hybridizing a second oligonucleotide to an adjacent region of the target nucleic acid;

b) contacting the hybridized first and second oligonucleotides with a cleavage enzyme to form a cleavage product; and

30 c) detecting the cleavage product by mass spectrometry as an indication of the presence of the target nucleic acid.

21. The method of claim 20 wherein the samples are from subjects in a healthy database.

22. The method of claim 18, wherein the polymorphism is identified by identifying target nucleic acids in a sample by primer oligo base extension  
5 (probe).

23. The method of 22, wherein primer oligo base extension, comprises:

- a) obtaining a nucleic acid molecule that contains a target nucleotide;
- b) optionally immobilizing the nucleic acid molecule onto a solid support,  
10 to produce an immobilized nucleic acid molecule;
- c) hybridizing the nucleic acid molecule with a primer oligonucleotide that is complementary to the nucleic acid molecule at a site adjacent to the target nucleotide;
- d) contacting the product of step c) with a composition comprising a  
15 dideoxynucleoside triphosphate or a 3'-deoxynucleoside triphosphates and a polymerase, so that only a dideoxynucleoside or 3'-deoxynucleoside triphosphate that is complementary to the target nucleotide is extended onto the primer; and
- e) detecting the extended primer, thereby identifying the target nucleotide.

20 24. The method of claim 23, wherein detection of the extended primer is effected by mass spectrometry, comprising:  
ionizing and volatilizing the product of step d) ; and  
detecting the extended primer by mass spectrometry, thereby identifying the target nucleotide.

25 25. The method of claim 24, wherein;  
samples are presented to the mass spectrometer as arrays on chips; and  
each sample occupies a volume that is about the size of the laser spot projected by the laser in a mass spectrometer used in matrix-assisted laser desorption/ionization (MALDI) spectrometry.

26. A combination, comprising:  
a database containing parameters associated with a datapoint  
representative of a subject from whom samples are obtained, wherein the  
subjects are healthy; and  
5 an indexed collection of the samples, wherein the index identifies the  
subject from whom the sample was obtained.

- 27 The combination of claim 26, wherein the parameter is selected  
from the group consisting of ethnicity, age, gender, height, weight, alcohol  
intake, number of pregnancies, number of live births, vegetarians, type of  
10 physical activity, state of residence and/or length of residence in a particular  
state, educational level, age of parent at death, cause of parent death, former or  
current smoker, length of time as a smoker, frequency of smoking, occurrence  
of disease in immediate family (parent, siblings, children), use of prescription  
drugs and/or reason therefor, length and/or number of hospital stays and  
15 exposure to environmental factors.

28. The combination of claim 26, wherein the database further  
contains genotypic data for each subject.

29. The combination of claim 26, wherein the samples are blood.

- 30 A data storage medium, comprising the database of claim 8.

- 20 31. A computer system, comprising the database of claim 8.

32. A system for high throughput processing of biological samples,  
comprising:

- 25 a process line comprising a plurality of processing stations, each of which  
performs a procedure on a biological sample contained in a  
reaction vessel;

- a robotic system that transports the reaction vessel from processing  
station to processing station;

- 30 a data analysis system that receives test results of the process line and  
automatically processes the test results to make a determination  
regarding the biological sample in the reaction vessel;

- a control system that determines when the test at each processing  
station is complete and, in response, moves the reaction vessel to

the next test station, and continuously processes reaction vessels one after another until the control system receives a stop instruction; and

a database of claim 8, wherein the samples tested by the automated process line comprise samples from subjects in the database.

33. The system of claim 32, wherein one of the processing stations comprises a mass spectrometer.

34. The system of claim 32, wherein the data analysis system processes the test results by receiving test data from the mass spectrometer such that the test data for a biological sample contains one or more signals, whereupon the data analysis system determines the area under the curve of each signal and normalizes the results thereof and obtains a substantially quantitative result representative of the relative amounts of components in the tested sample.

35. A method for high throughput processing of biological samples, the method comprising:

transporting a reaction vessel along a system of claim 32, comprising a process line having a plurality of processing stations, each of which performs a procedure on one or more biological samples contained in the reaction vessel;

determining when the test procedure at each processing station is complete and, in response, moving the reaction vessel to the next processing station;

receiving test results of the process line and automatically processing the test results to make a data analysis determination regarding the biological samples in the reaction vessel; and

processing reaction vessels continuously one after another until receiving a stop instruction, wherein the samples tested by the automated process line comprise samples from subjects in the database.

36. The method of 35, wherein one of the processing stations comprises a mass spectrometer.

37. The method of claim 36, wherein the samples are analyzed by a method comprising primer oligo base extension (probe).

38. The method of claim 37, further comprising:

processing the test results by receiving test data from the mass  
5 spectrometer such that the test data for a biological sample contains one or more signals or numerical values representative of signals, whereupon the data analysis system determines the area under the curve of each signal and normalizes the results thereof and obtains a substantially quantitative result representative of the relative amounts of components in the tested sample.

10 39. The method of claim 37, wherein primer oligo base extension, comprises:

a) obtaining a nucleic acid molecule that contains a target nucleotide;

b) optionally immobilizing the nucleic acid molecule onto a solid support, to produce an immobilized nucleic acid molecule;

15 c) hybridizing the nucleic acid molecule with a primer oligonucleotide that is complementary to the nucleic acid molecule at a site adjacent to the target nucleotide;

d) contacting the product of step c) with composition comprising a dideoxynucleoside triphosphate or a 3'-deoxynucleoside triphosphates and a  
20 polymerase, so that only a dideoxynucleoside or 3'-deoxynucleoside triphosphate that is complementary to the target nucleotide is extended onto the primer; and  
e) detecting the primer, thereby identifying the target nucleotide.

40. The method of 39, wherein detection of the extended primer is effected by mass spectrometry, comprising:

25 ionizing and volatilizing the product of step d); and

detecting the extended primer by mass spectrometry, thereby identifying the target nucleotide.

41. The method of claim 36, wherein the target nucleic acids in the sample are detected and/or identified by a method, comprising the steps of:

30 a) hybridizing a first oligonucleotide to the target nucleic acid;

b) hybridizing a second oligonucleotide to an adjacent region of the target nucleic acid;

c) ligating then hybridized oligonucleotides; and

c) detecting hybridized first oligonucleotide by mass spectrometry as an indication of the presence of the target nucleic acid.

42. The method of claim 36, wherein the target nucleic acids in the sample are detected and/or identified by a method, comprising the steps of:

a) hybridizing a first oligonucleotide to the target nucleic acid and hybridizing a second oligonucleotide to an adjacent region of the target nucleic acid;

b) contacting the hybridized first and second oligonucleotides with a cleavage enzyme to form a cleavage product; and

c) detecting the cleavage product by mass spectrometry as an indication of the presence of the target nucleic acid.

43. A method of producing a database stored in a computer memory, comprising:

identifying healthy members of a population;

obtaining identifying and historical information and data relating to the identified members of the population;

entering the member-related data into the computer memory database for each identified member of the population and associating the member and the data with an indexer.

44. The method of claim 43, further comprising:

obtaining a body tissue or body fluid sample of an identified member;

analyzing the body tissue or body fluid in the sample; and

entering the results of the analysis for each member into the computer memory database and associating each result with the indexer representative of each member.

45. A database produced by the method of claim 43.

46. A database produced by the method of claim 44.

47. The database of claim 8, wherein:

the organisms are selected from among animals, bacteria, fungi, protozoans and parasites and

each datapoint is associated with parameters representative of the organism type and identifying information.

48. The database of claim 43, further comprising, phenotypic data regarding each subject.

5 49. The database of claim 47 that is a relational database and the parameters are the answers to the questions in the questionnaire.

50. The database of claim 8, further comprising, genotypic data of nucleic acid of the subject, wherein genotypic data includes, but is not limited to, genetic markers, non-coding regions, 10 microsatellites, restriction fragment length polymorphisms (RFLPs), variable number tandem repeats (VNTRs), historical day of the organism, the medical history of the subject, phenotypic information, and other information.

51. A database, comprising data records stored in computer memory, wherein the data records contain information that identifies healthy members of 15 a population, and also contain identifying and historical information and data relating to the identified members.

52. The database of claim 51, further comprising an index value for each identified member that associates each member of the population with the identifying and historical information and data.

20 53. A computer system, comprising the database of claim 51.

54. An automated process line, comprising the database of claim 51.

55. A method for determining a polymorphism that correlates with age, ethnicity or gender, comprising:

25 identifying a polymorphism; and determining the frequency of the polymorphism with increasing age, with ethnicity or with gender in a healthy population.

56. A method for determining whether a polymorphism correlates with susceptibility to morbidity, early mortality, or morbidity and early mortality, comprising;

30 identifying a polymorphism; and

determining the frequency of the polymorphism with increasing age in a healthy population.

57. A high throughput method of determining frequencies of genetic variations, comprising:

selecting a healthy target population and a genetic variation to be assessed;

5 pooling a plurality of samples of biopolymers obtained from members of the population,

determining or detecting the biopolymer that comprises the variation by mass spectrometry;

obtaining a mass spectrum or a digital representation thereof; and

10 determining the frequency of the variation in the population.

58. The method of claim 57, wherein:

the variation is selected from the group consisting of an allelic variation, a post-translational modification, a nucleic modification, a label, a mass modification of a nucleic acid and methylation; and/or

15 the biopolymer is a nucleic acid, a protein, a polysaccharide, a lipid, a small organic metabolite or intermediate, wherein the concentration of biopolymer of interest is the same in each of the samples; and/or

the frequency is determined by assessing the method comprising determining the area under the peak in the mass spectrum or digital  
20 representation thereof corresponding to the mass of the biopolymer comprising the genomic variation.

59. The method of claim 58, wherein the method for determining the frequency is effected by determining the ratio of the signal or the digital representation thereof to the total area of the entire mass spectrum, which is  
25 corrected for background.

60. A method for discovery of a polymorphism in a population, comprising:

sorting the database of claim 8 according to a selected parameter to identify samples that match the selected parameter;

30 isolating a biopolymer from each identified sample;

optionally pooling each isolated biopolymer;

optionally amplifying the amount of biopolymer;

cleaving the pooled biopolymers to produce fragments thereof;

obtaining a mass spectrum of the resulting fragments and comparing the mass spectrum with a control mass spectrum to identify differences between the spectra and thereby identifying any polymorphisms; wherein:

5       the control mass spectrum is obtained from unsorted samples in the collection or samples sorted according to a different parameter.

61. The method of claim 60, wherein cleaving is effected by contacting the biopolymer with an enzyme.

62. The method of claim 61, wherein the enzyme is selected from the  
10       group consisting of nucleotide glycosylase, a nickase and a type IIS restriction enzyme.

63. The method of claim 60, wherein the biopolymer is a nucleic acid or a protein.

64. The method of claim 60, wherein the the mass spectrometric  
15       format is selected from among Matrix-Assisted Laser Desorption/Ionization, Time-of-Flight (MALDI-TOF), Electrospray (ES), IR-MALDI, Ion Cyclotron Resonance (ICR), Fourier Transform and combinations thereof.

65. A method for discovery of a polymorphism in a population, comprising:

20       obtaining samples of body tissue or fluid from a plurality of organisms;  
          isolating a biopolymer from each sample;  
          pooling each isolated biopolymer;  
          optionally amplifying the amount of biopolymer;  
          cleaving the pooled biopolymers to produce fragments thereof;  
25       obtaining a mass spectrum of the resulting fragments;  
          comparing the frequency of each fragment to identify fragments present in amounts lower than the average frequency, thereby identifying any polymorphisms.

66. The method of claim 65, wherein cleaving is effected by contacting  
30       the biopolymer with an enzyme.

67. The method of claim 66, wherein the enzyme is selected from the group consisting of nucleotide glycosylase, a nickase and a type IIS restriction enzyme.

68. The method of claim 65, wherein the biopolymer is a nucleic acid  
5 or a protein.

69. The method of claim 65, wherein the the mass spectrometric format is selected from among Matrix-Assisted Laser Desorption/Ionization, Time-of-Flight (MALDI-TOF), Electrospray (ES), IR-MALDI, Ion Cyclotron Resonance (ICR), Fourier Transform and combinations thereof.

70. The method of claim 65, wherein the samples are obtained from  
10 healthy subjects.

71. A method of correlating a polymorphism with a parameter, comprising:

15 sorting the database of claim 8 according to a selected parameter to

identify samples that match the selected parameter;

isolating a biopolymer from each identified sample;

pooling each isolated biopolymer;

optionally amplifying the amount of biopolymer;

20 determining the frequency of the polymorphism in the pooled  
biopolymers, wherein:

an alteration of the frequency of the polymorphism compared to a control, indicates a correlation of the polymorphism with the selected parameter; and

the control is the frequency of the polymorphism in pooled biopolymers obtained from samples identified from an unsorted database or from a database  
25 sorting according to a different parameter.

72. The method claim 71, wherein the parameter is selected from the group consisting of ethnicity, age, gender, height, weight, alcohol intake, number of pregnancies, number of live births, vegetarians, type of physical activity, state of residence and/or length of residence in a particular state,  
30 educational level, age of parent at death, cause of parent death, former or current smoker, length of time as a smoker, frequency of smoking, occurrence of a disease in immediate family (parent, siblings, children), use of prescription

drugs and/or reason therefor, length and/or number of hospital stays and exposure to environmental factors.

73. The method claim 72, wherein the parameter is occurrence of disease or a particular disease in an immediate family member, thereby  
5 correlating the polymorphism with the disease.

74. The method of claim 71, wherein the pooled biopolymers are pooled nucleic acid molecules.

75. The method of claim 74, wherein the polymorphism is detected by primer oligo base extension (PROBE).

10 76. The method of 75, wherein primer oligo base extension, comprises:

a) optionally immobilizing the nucleic acid molecules onto a solid support, to produce immobilized nucleic acid molecules;

b) hybridizing the nucleic acid molecules with a primer oligonucleotide  
15 that is complementary to the nucleic acid molecule at a site adjacent to the polymorphism;

c) contacting the product of step c) with composition comprising a dideoxynucleoside triphosphate or a 3'-deoxynucleoside triphosphates and a polymerase, so that only a dideoxynucleoside or 3'-deoxynucleoside triphosphate  
20 that is complementary to the polymorphism is extended onto the primer; and

d) detecting the extended primer, thereby detecting the polymorphism in nucleic acid molecules in the pooled nucleic acids.

77. The method of claim 76, wherein detecting is effected by mass spectrometry.

25 78. The method of claim 71, wherein the frequency is percentage of nucleic acid molecules in the pooled nucleic acids that contain the polymorphism.

79. The method of claim 78, wherein the ratio is determined by obtaining mass spectra of the pooled nucleic acids.

30 80. The method of claim 72, wherein the parameter is age, thereby correlating the polymorphism with suceptibility to morbidity, early mortality or morbidity and early mortality.

81. A method for haplotyping polymorphisms in a nucleic acid, comprising:

- (a) sorting the database of claim 8 according to a selected parameter to identify samples that match the selected parameter;
- 5 (b) isolating nucleic acid from each identified sample;
- (c) optionally pooling each isolated nucleic acid;
- (d) amplifying the amount of nucleic acid;
- (e) forming single-stranded nucleic acid and splitting each single-strand into a separate reaction vessel;
- 10 (f) contacting each single-stranded nucleic acid with an adaptor nucleic acid to form an adaptor complex;
- (g) contacting the adaptor complex with a nuclease and a ligase;
- (h) contacting the products of step (g) with a mixture that is capable of amplifying a ligated adaptor to produce an extended product;
- 15 (i) obtaining a mass spectrum of each nucleic acid resulting from step (h) and detecting a polymorphism by identifying a signal corresponding to the extended product;
- (j) repeating steps (f) through (i) utilizing an adaptor nucleic acid able to hybridize with another adapter nucleic acid that hybridizes to a different
- 20 sequence on the same strand; whereby  
the polymorphisms are haplotyped by detecting more than one extended product.

82. The method of claim 1, wherein the nuclease is Fen-1.

83. A method for haplotyping polymorphisms in a population, comprising:

- sorting the database of claim 8 according to a selected parameter to identify samples that match the selected parameter;
- isolating a nucleic acid from each identified sample;
- pooling each isolated nucleic acid;
- 30 optionally amplifying the amount of nucleic acid;
- contacting the nucleic acid with at least one enzyme to produce fragments thereof;

obtaining a mass spectrum of the resulting fragments; whereby:  
the polymorphisms are detected by detecting signals corresponding to the  
polymorphisms; and

the polymorphisms are haplotyped by determining from the mass  
5 spectrum that the polymorphisms are located on the same strand of the nucleic  
acid.

84. The method of claim 83, wherein the enzyme is a nickase.

85. The method of claim 84, wherein the nickase is selected from the  
group consisting of NY2A and NYS1.

10 86. A method for detecting methylated nucleotides within a nucleic  
acid sample, comprising:

splitting a nucleic acid sample into separate reaction vessels;

contacting nucleic acid in one reaction vessel with bisulfite;

amplifying the nucleic acid in each reaction vessel;

15 cleaving the nucleic acids in each reaction vessel to produce fragments  
thereof;

obtaining a mass spectrum of the resulting fragments from one reaction  
vessel and another mass spectrum of the resulting fragments from another  
reaction vessel; whereby:

20 cytosine methylation is detected by identifying a difference in signals  
between the mass spectra.

87. The method of claim 86, wherein:

the step of amplifying is carried out in the presence of uracil; and

the step of cleaving is effected by a uracil glycosylase.

25 88. A method for identifying a biological sample, comprising:  
generating a data set indicative of the composition of the biological  
sample;

denoising the data set to generate denoised data;

deleting the baseline from the denoised data to generate an intermediate

30 data  
set;

defining putative peaks for the biological sample;

using the putative peaks to generate a residual baseline;  
removing the residual baseline from the intermediate data set to generate  
a corrected data set;

5        locating, responsive to removing the residual baseline, a probable peak in  
the  
corrected data set; and

identifying, using the located probable peak, the biological sample;  
wherein the generated biological sample data set comprises data from  
sense  
10        strands and antisense strands of assay fragments.

89.     The method according to claim 88, wherein identifying includes  
combining  
data from the sense strands and the antisense strands, and comparing the data  
against expected sense strand and antisense strand values, to identify the  
15        biological  
sample.

90.     The method according to claim 88, wherein identifying includes  
deriving a peak probability for the probable peak, in accordance with whether the  
probable peak is from sense strand data or from antisense strand data.

20        91.     The method according to claim 88, wherein identifying includes  
deriving a peak probability for the probable peak and applying an allelic penalty in  
response to a  
ratio between a calculated area under the probable peak and a calculated  
expected average area under all peaks in the data set.

25        92.     A method for identifying a biological sample, comprising:  
generating a data set indicative of the composition of the biological  
sample;  
denoising the data set to generate denoised data;  
deleting the baseline from the denoised data to generate an intermediate  
30        data  
set;

defining putative peaks for the biological sample; using the  
putative peaks to generate a residual baseline;  
removing the residual baseline from the intermediate data set to generate  
a

5 corrected data set;

locating, responsive to removing the residual baseline, a probable peak in  
the corrected data set; and

identifying, using the located probable peak, the biological sample;

wherein identifying includes deriving a peak probability for the probable  
10 peak and

applying an allelic penalty in response to a ratio between a calculated  
area under the  
probable peak and a calculated expected average area under all peaks in the data  
set.

15 93. The method according to claim 92, wherein identifying includes  
comparing

data from probable peaks that did not receive an applied allelic penalty to  
determine their mass in accordance with oligonucleotide biological data.

94. The method according to claim 92, wherein the allelic penalty is  
20 not applied to probable peaks whose ratio of area under the peak to the  
expected area value is greater than 30%.

95. A method for detecting a polymorphism in a nucleic acid,  
comprising:

25 amplifying a region of the nucleic acid to produce an amplicon, wherein  
the resulting amplicon comprises one or more enzyme restriction sites;

contacting the amplicon with a restriction enzyme to produce fragments;

obtaining a mass spectrum of the resulting fragments and analyzing  
signals in the mass spectrum by the method of claim 88; whereby:

the polymorphism is detected from the pattern of the signals.

30 96. A subcollection of samples from a target population, comprising:  
a plurality of samples, wherein the samples are selected from the group  
consisting of nucleic acids, fetal tissue, protein samples; and

a symbology on the containers containing the samples, wherein the symbology is representative of the source and/or history of each sample, wherein:

the target population is a healthy population that has not been selected  
5 for any disease state;

the collection comprises samples from the healthy population; and

the subcollection is obtained by sorting the collection according to specified parameters.

97. The combination of claim 26, wherein the samples are selected  
10 selected from the group consisting of nucleic acids, fetal tissue, protein, tissue, body fluid, cell, seed, microbe, pathogen and reproductive tissue samples.

98. A combination, comprising the database of claim 8 and a mass spectrometer.

99. The combination of claim 98 that is an automated process line for  
15 analyzing biological samples.

100. A system for high throughput processing of biological samples, comprising:

a database of claim 8, wherein the samples tested by the automated  
process line comprise samples from subjects in the database; and  
20 a mass spectrometry for analysis of biopolymers in the samples.

## DNA Bank

Number of Samples	3912
-------------------	------

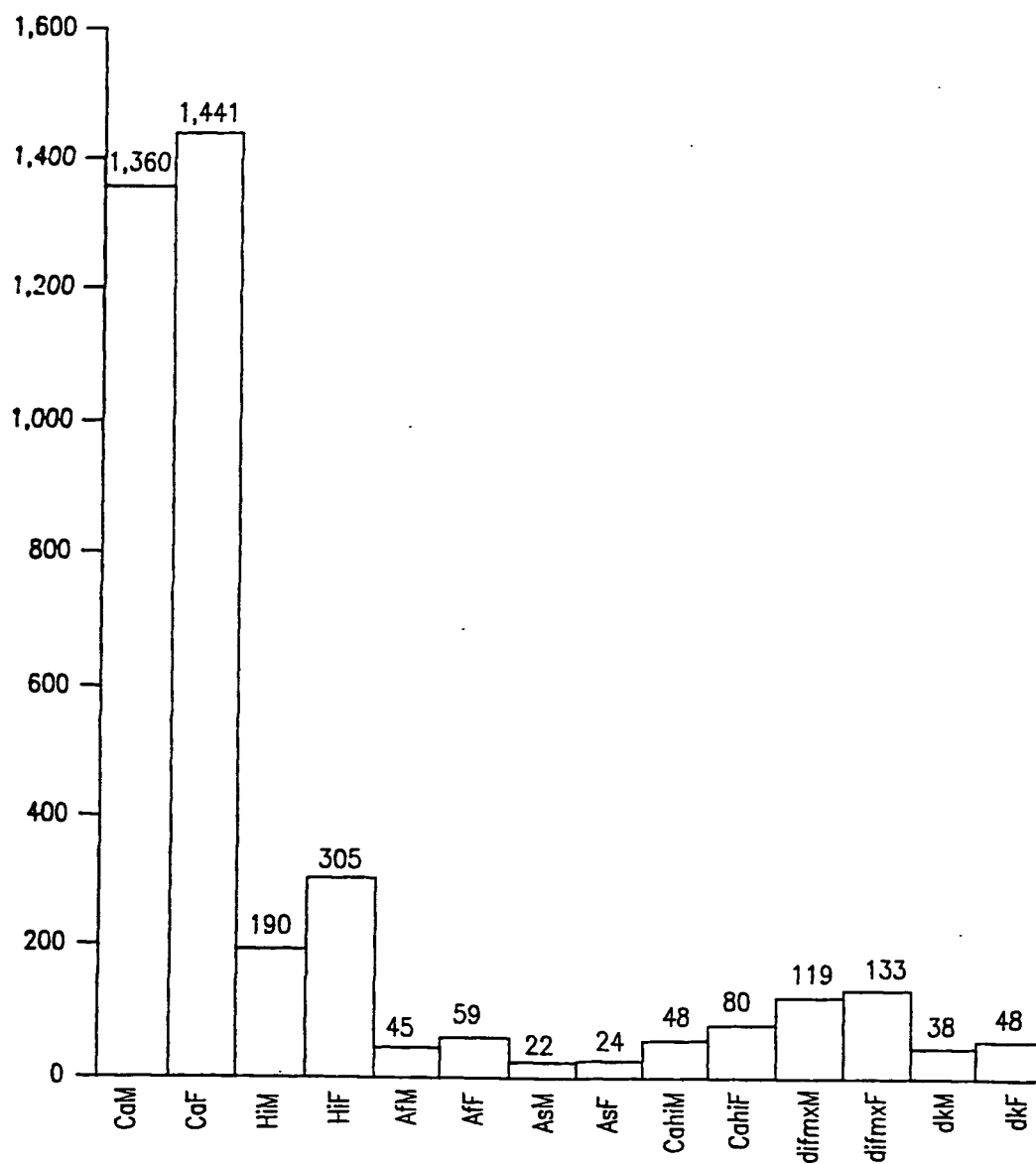


FIG. 1A

## Caucasians

Number of Samples	2801
-------------------	------

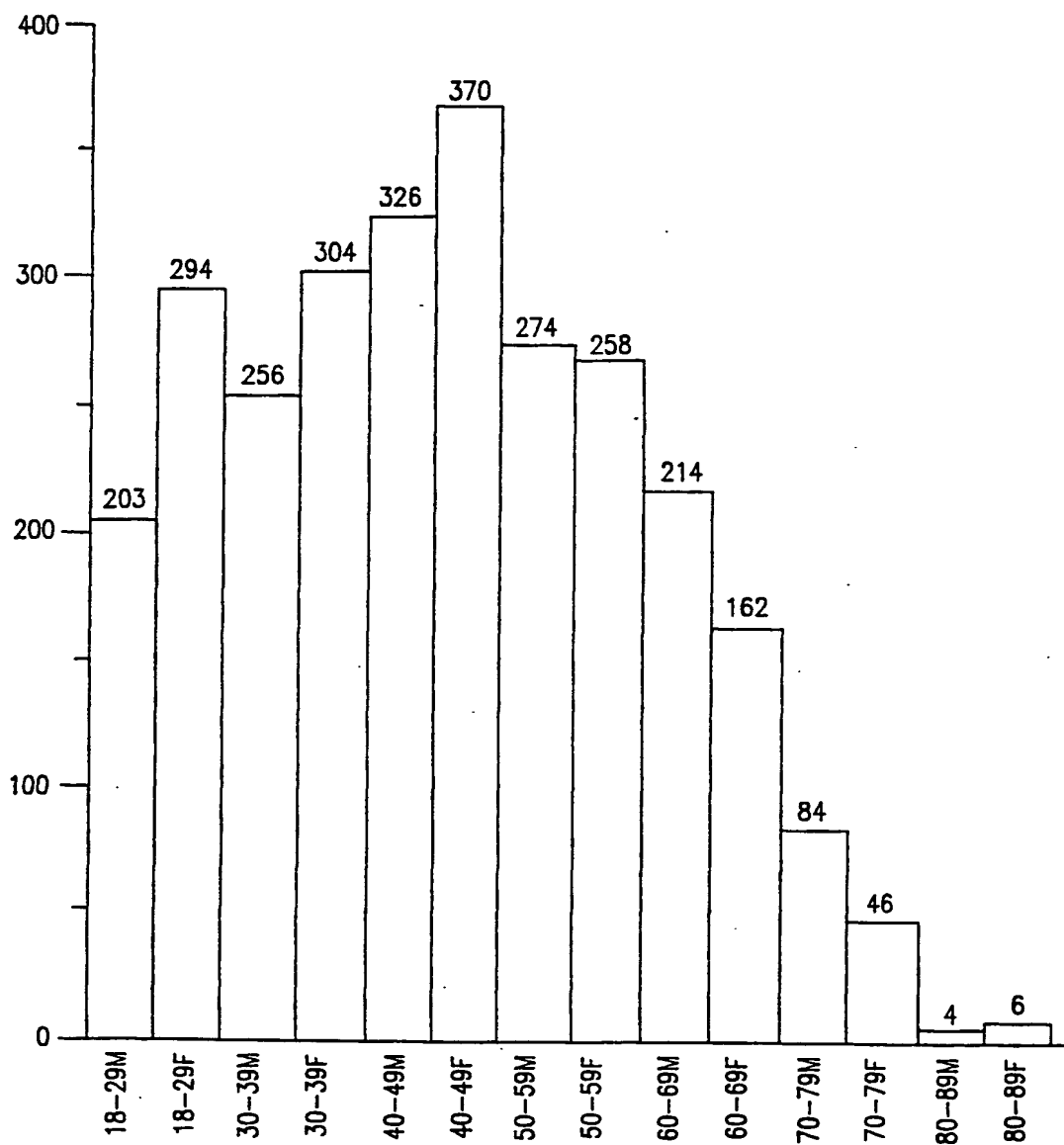


FIG. 1B

## Hispanics

Number of Samples	495
-------------------	-----

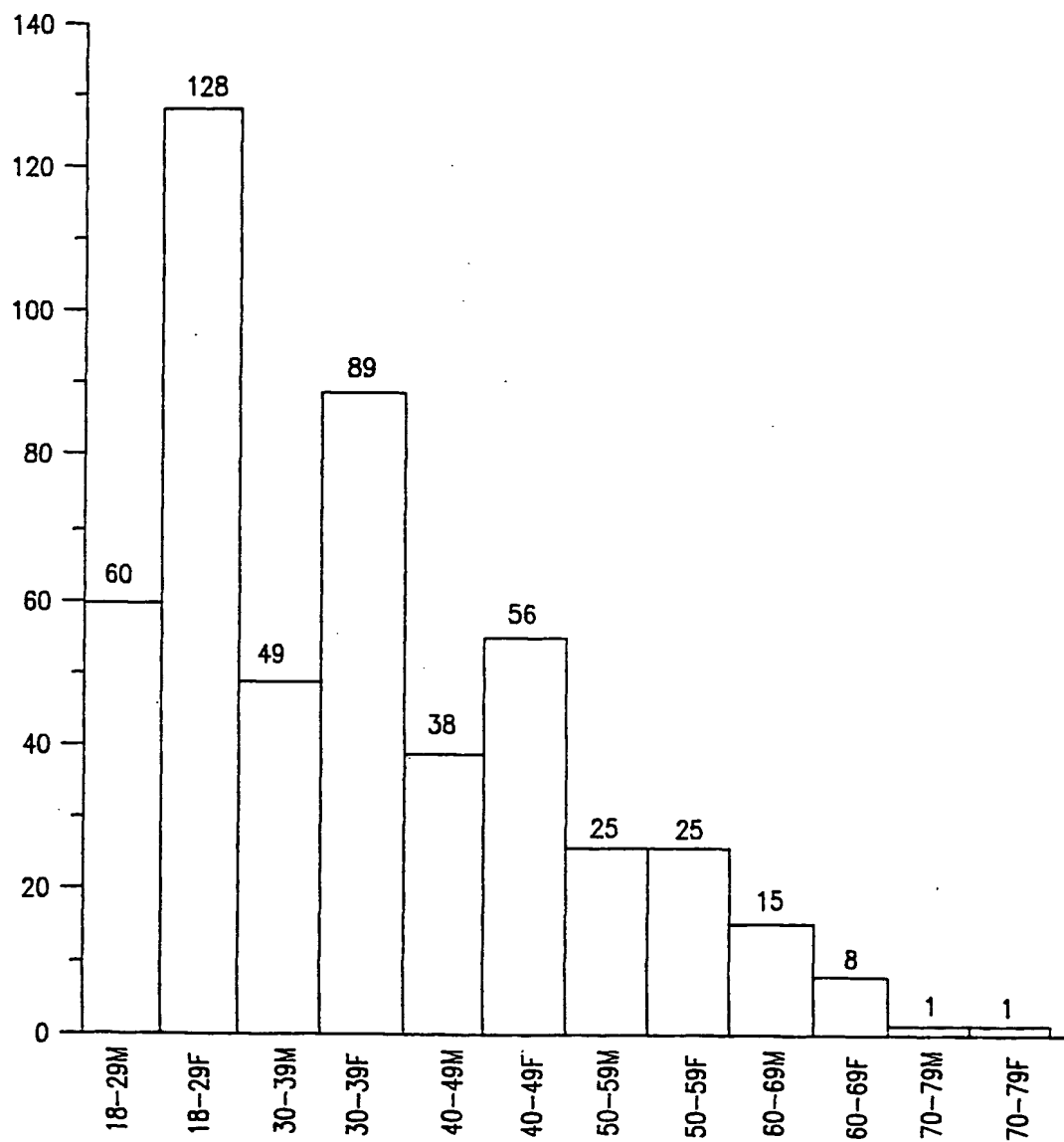
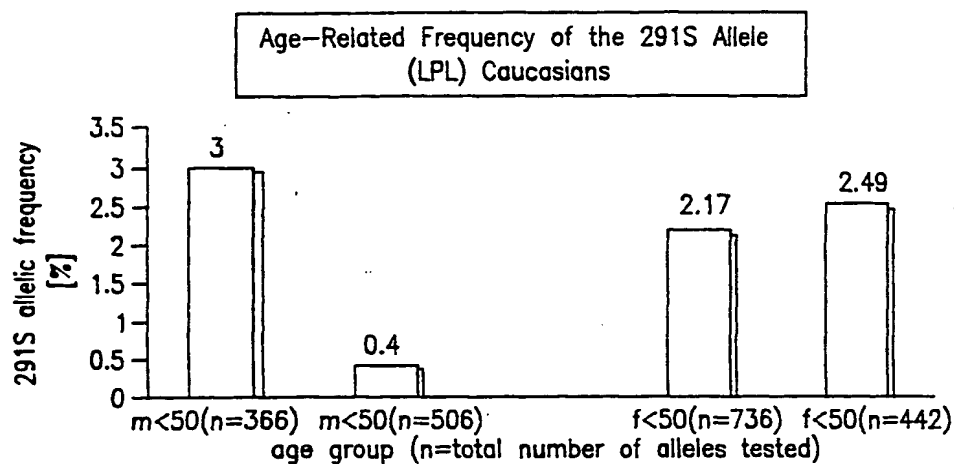


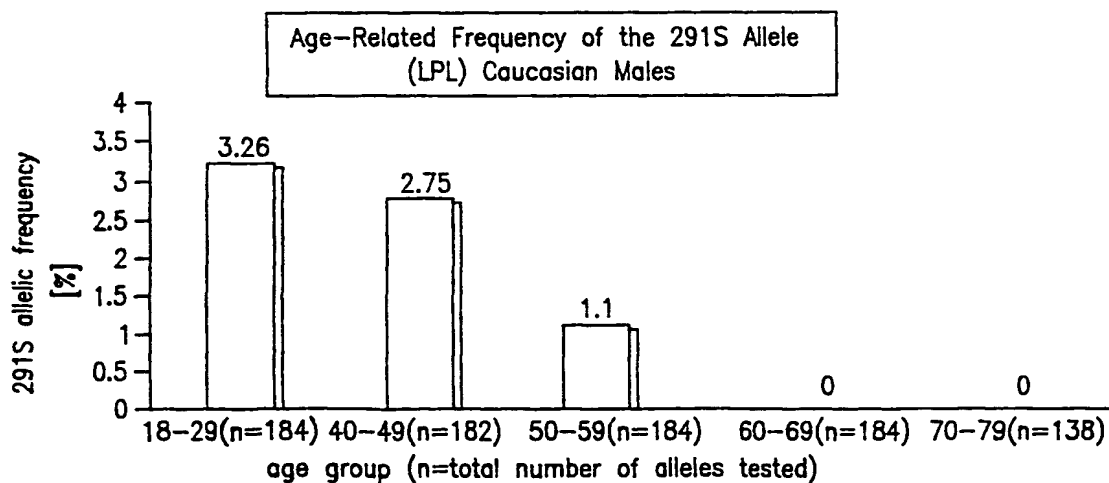
FIG. 1C

SUBSTITUTE SHEET (RULE 26)



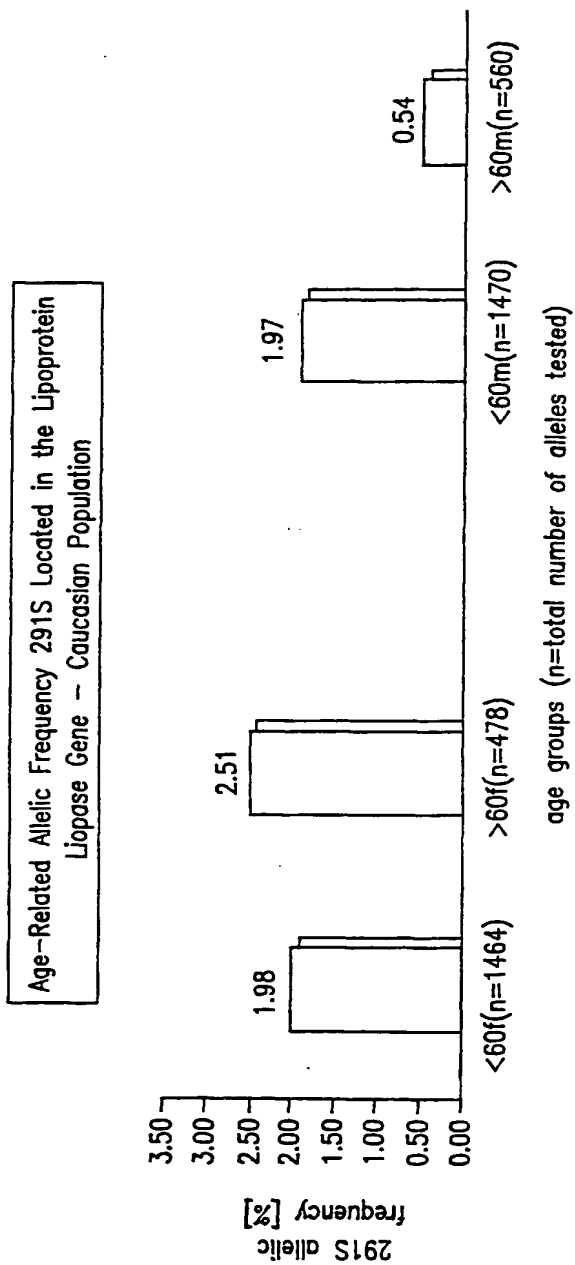
age- and sex-distribution of the 291S allele of the lipoprotein lipase gene. A total of 436 males and 586 females were investigated.

FIG. 2A



Age-related distribution of the 291S allele of the lipoprotein lipase gene within the male Caucasian population. A total of 436 males were tested.

FIG. 2B



Significance: Allele frequency of Ser291 drops from 1.97% to 0.54% in males;  $p=0.009$

FIG. 2C

Questionnaire for  
Population-Based  
Sample Banking

Data Collection Form

Collection Information

Consent Form Signed Yes No

Date of Collection (MM/DD/YY)\_\_\_/\_\_\_/98

Time of Sample Collection(nearest hour in 24 hour clock format)\_\_\_\_\_

Initials of Data Collector\_\_\_\_\_Collecting Agency\_\_\_\_\_

(DO NOT COMPLETE: (For Date Entry Only)Sample\_\_\_\_\_intact\_\_\_\_\_lost\_\_\_\_\_broken

Affix Barcode Here

Donor information

Sex: ☐ Male ☐ Female

Date of Birth (MM/YY)\_\_\_/\_\_\_

In which state do you live? \_\_\_\_\_

How long have you lived there ? \_\_\_\_\_ Years

What is your highest grade you completed in school?

☐ less than 8th grade ☐ 8th,9th,10th or 11th grade ☐ high school graduate or equivalency

☐ some college 2 yr degree ☐ college graduate 4 yr degree ☐ post graduate education or degree

To the best of your knowledge what is the Ethnic Origin of your:

Father

Mother

- |                          |                          |   |
|--------------------------|--------------------------|---|
| <input type="checkbox"/> | <input type="checkbox"/> | Caucasian (please check specific geographic area below if known)                                    |
| <input type="checkbox"/> | <input type="checkbox"/> | Northern Europe (Austria,Denmark,Finland,France,Germany,Netherlands,Norway,Sweden,Switzerland,U.K.) |
| <input type="checkbox"/> | <input type="checkbox"/> | Southern Europe (Greece,Italy,Spain)  |
| <input type="checkbox"/> | <input type="checkbox"/> | Eastern Europe (Czechoslovakia,Hungary,Poland,Russia,Yugoslavia)                                    |
| <input type="checkbox"/> | <input type="checkbox"/> | Middle Eastern (Israel,Egypt,Iran,Iraq,Jordan,Syria, other Arab States)                             |
| <input type="checkbox"/> | <input type="checkbox"/> | African-American  |
| <input type="checkbox"/> | <input type="checkbox"/> | Hispanic (please check specific geographic area below if known)                                     |
| <input type="checkbox"/> | <input type="checkbox"/> | Mexico  |
| <input type="checkbox"/> | <input type="checkbox"/> | Central America, South American   |
| <input type="checkbox"/> | <input type="checkbox"/> | Cuba, Puerto Rico, other Caribbean  |
| <input type="checkbox"/> | <input type="checkbox"/> | Asian (please check specific geographic area below if known)  |
| <input type="checkbox"/> | <input type="checkbox"/> | Japanese  |
| <input type="checkbox"/> | <input type="checkbox"/> | Chinese   |
| <input type="checkbox"/> | <input type="checkbox"/> | Korean  |
| <input type="checkbox"/> | <input type="checkbox"/> | Vietnamese  |
| <input type="checkbox"/> | <input type="checkbox"/> | other Asian   |
| <input type="checkbox"/> | <input type="checkbox"/> | Other _____   |
| <input type="checkbox"/> | <input type="checkbox"/> | Don't know  |

Health information: Have you or has anyone in your immediate family(parents,brothers,sisters, or your children) had the following? Check all that apply

Disease:

You Mother Father Sister Brother Child

Heart Disease Stroke or Arteriosclerosis
Cancer (Specify type if known)
Alzheimer's Disease or Dementia
Chronic inflammatory or Autoimmune Disease
Nervous System Disease like Multiple Sclerosis
Other (please specify)

Additional health information details you would like to provide:

FIG. 3

SUBSTITUTE SHEET (RULE 26)

# Sample Banks

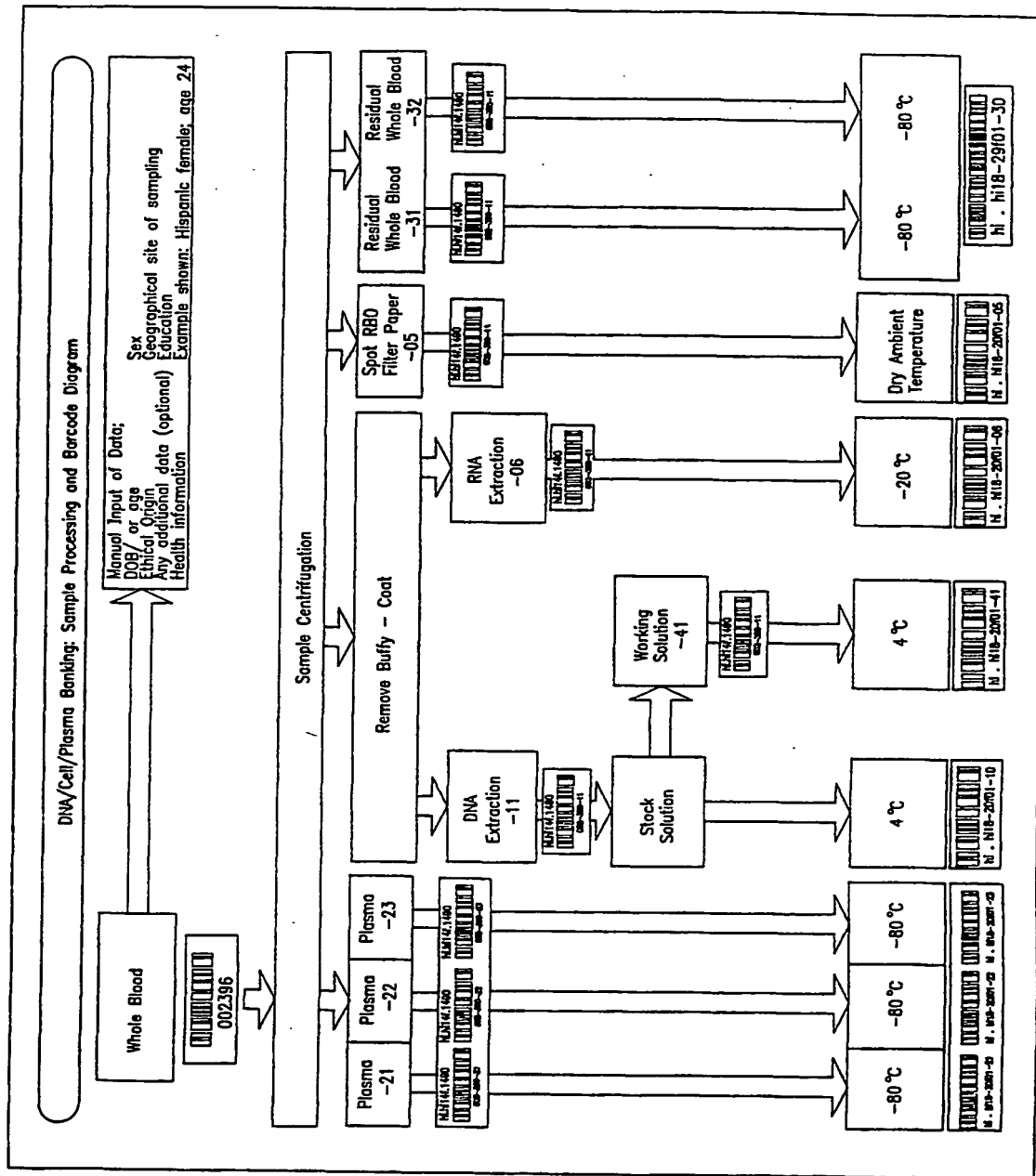


FIG. 4

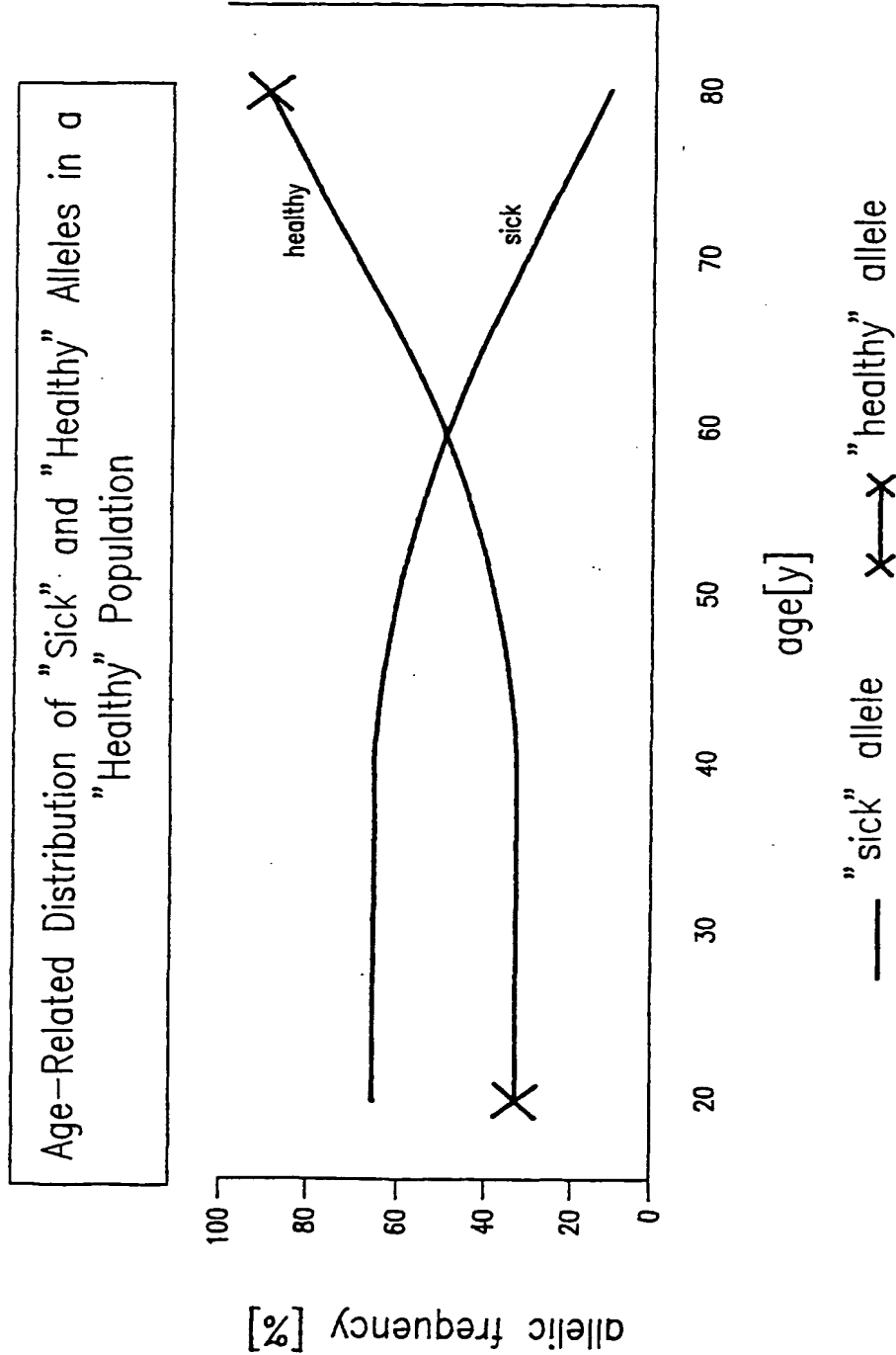


FIG. 5

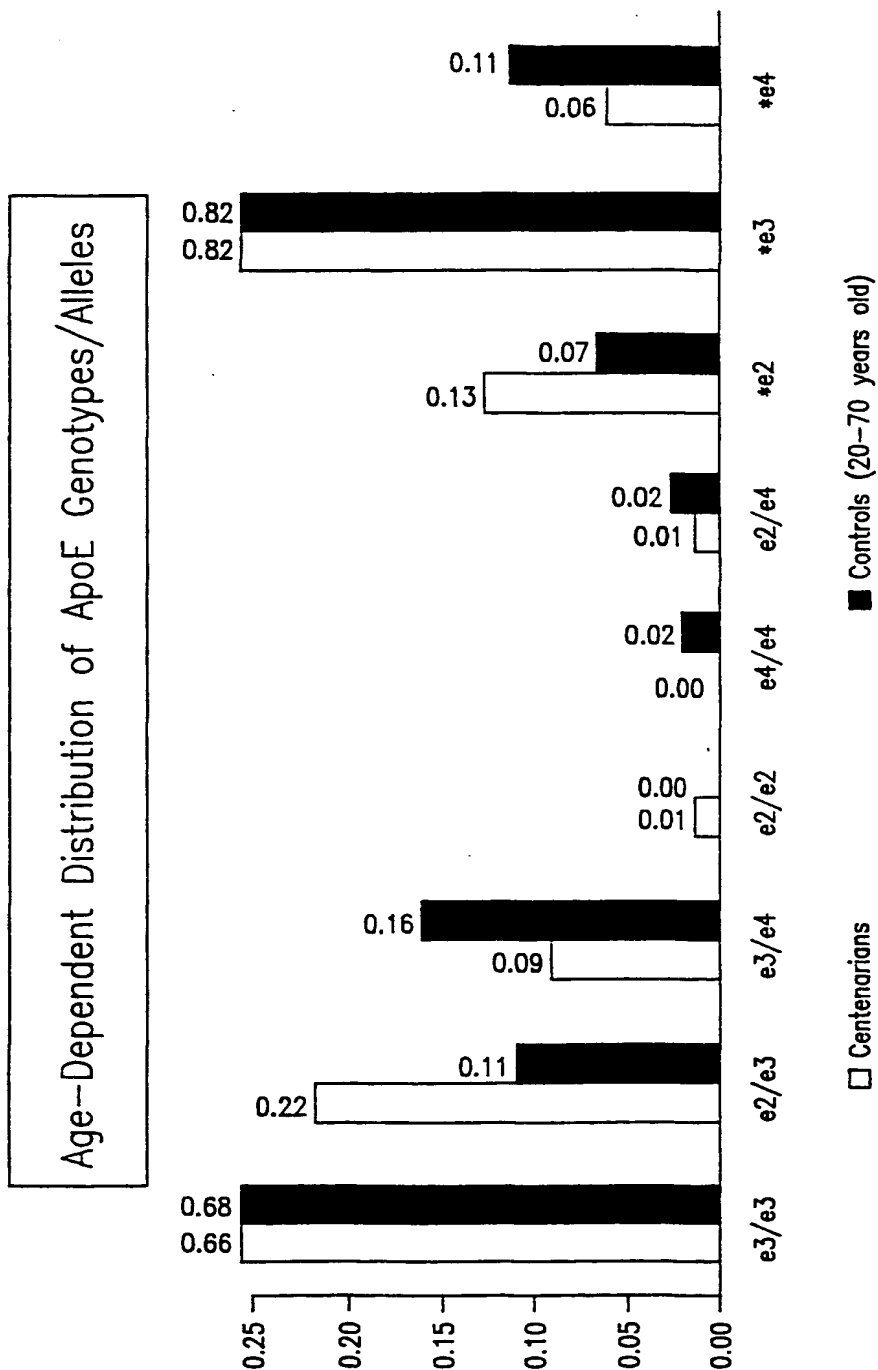


FIG. 6

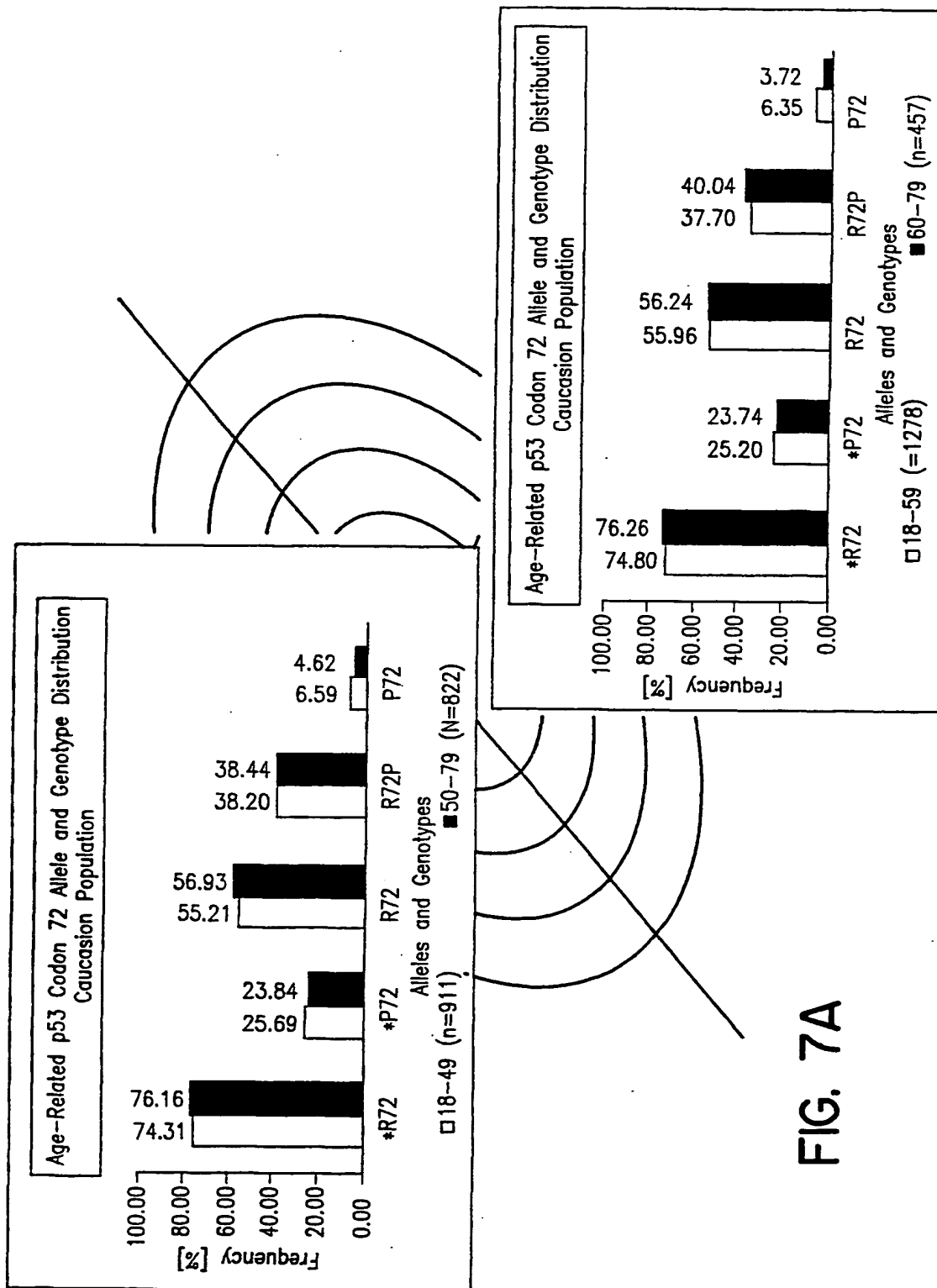


FIG. 7A

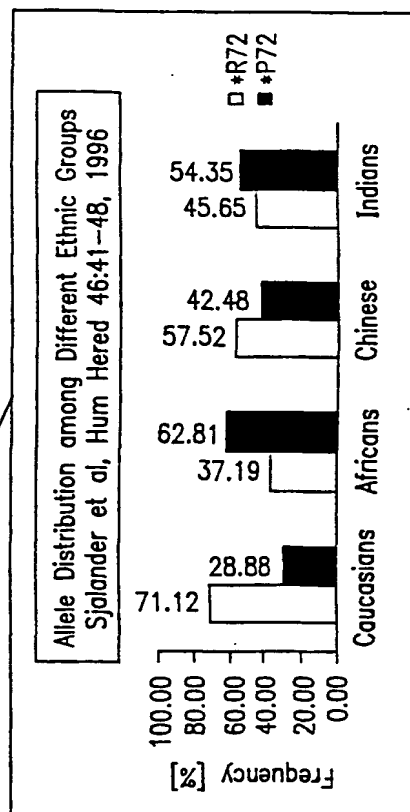
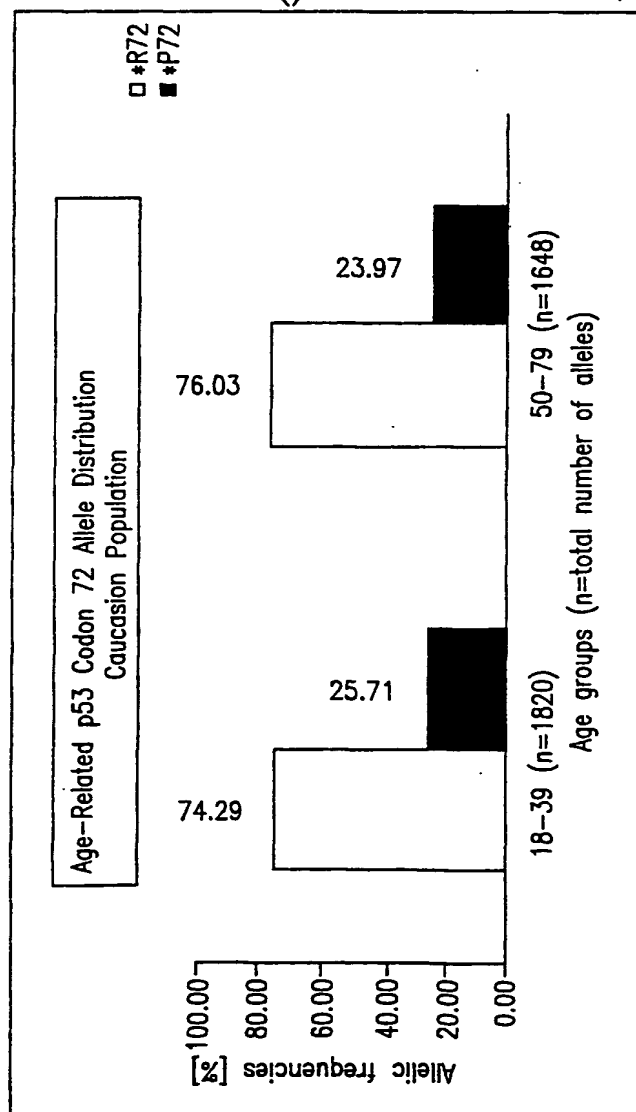
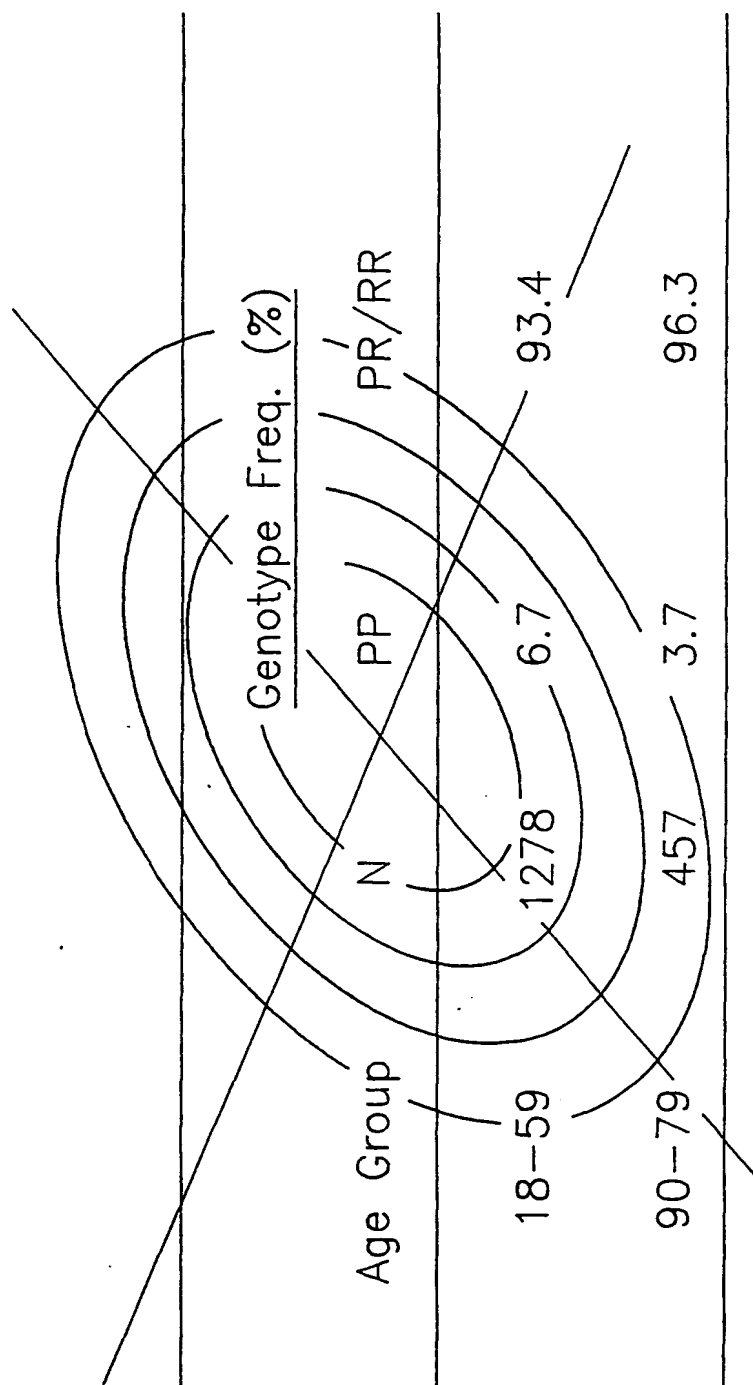


FIG. 7B

FIG. 7C

P53 PP vs. PR/RR Genotype Distribution  
By Age cut point = 59



Sample Size : 1735

$\chi^2$ : 5.2 (1 d.f.),  $P = 0.02$

Genomic Organization of the p53 Gene

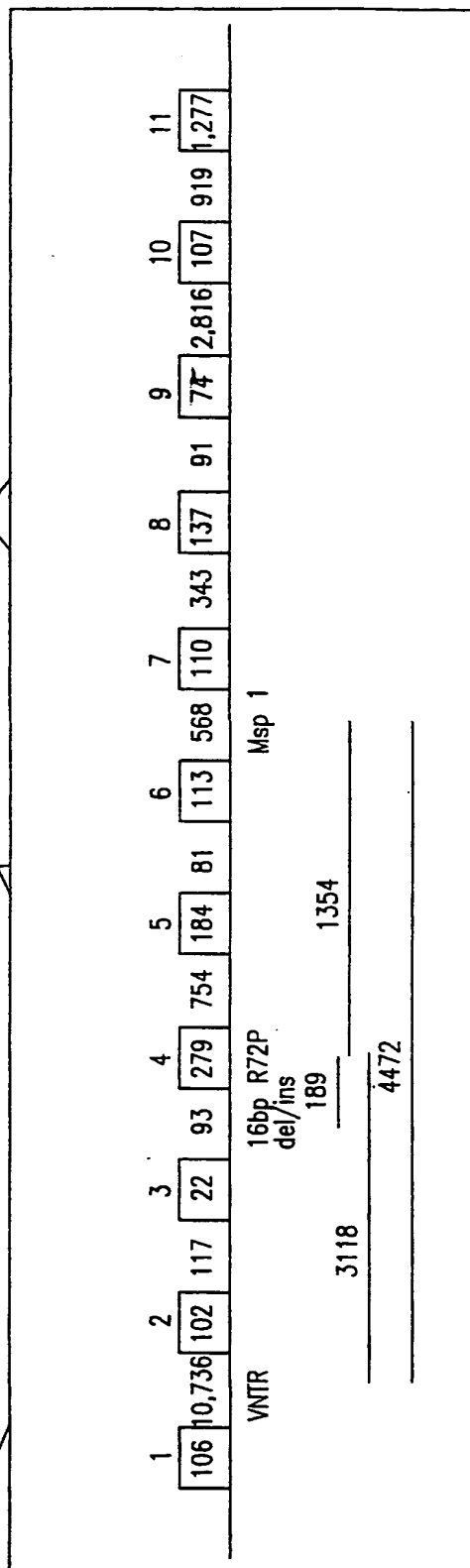
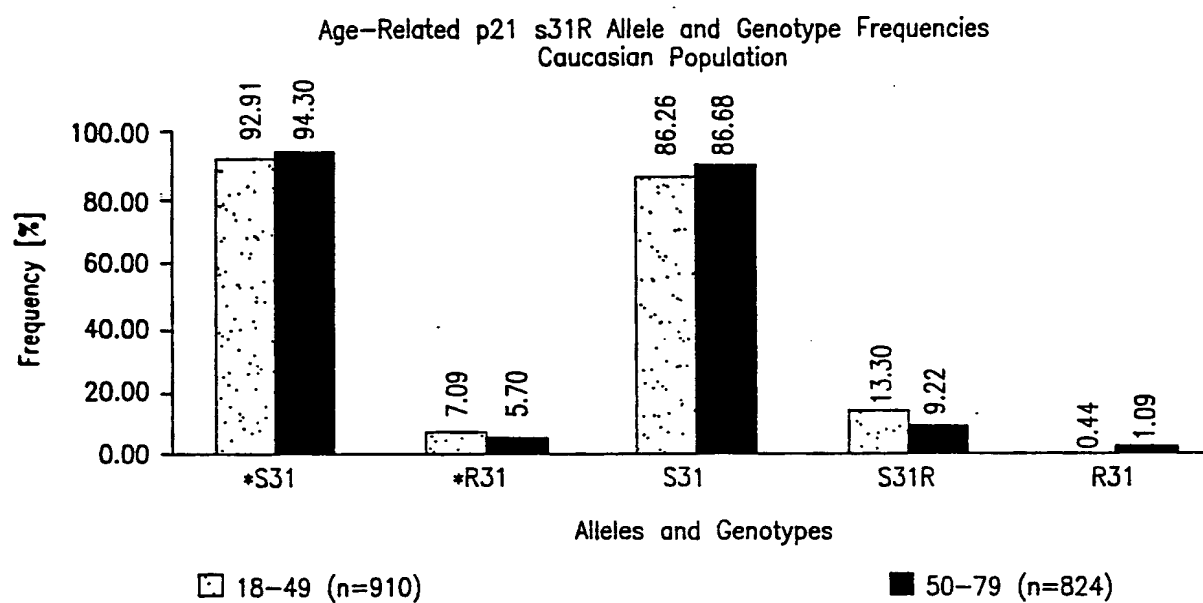


FIG. 7D



Significance: Genotype frequency of SR heterozygous  
drops from 13.3% to 9.2%;  $p=0.009$

FIG. 8

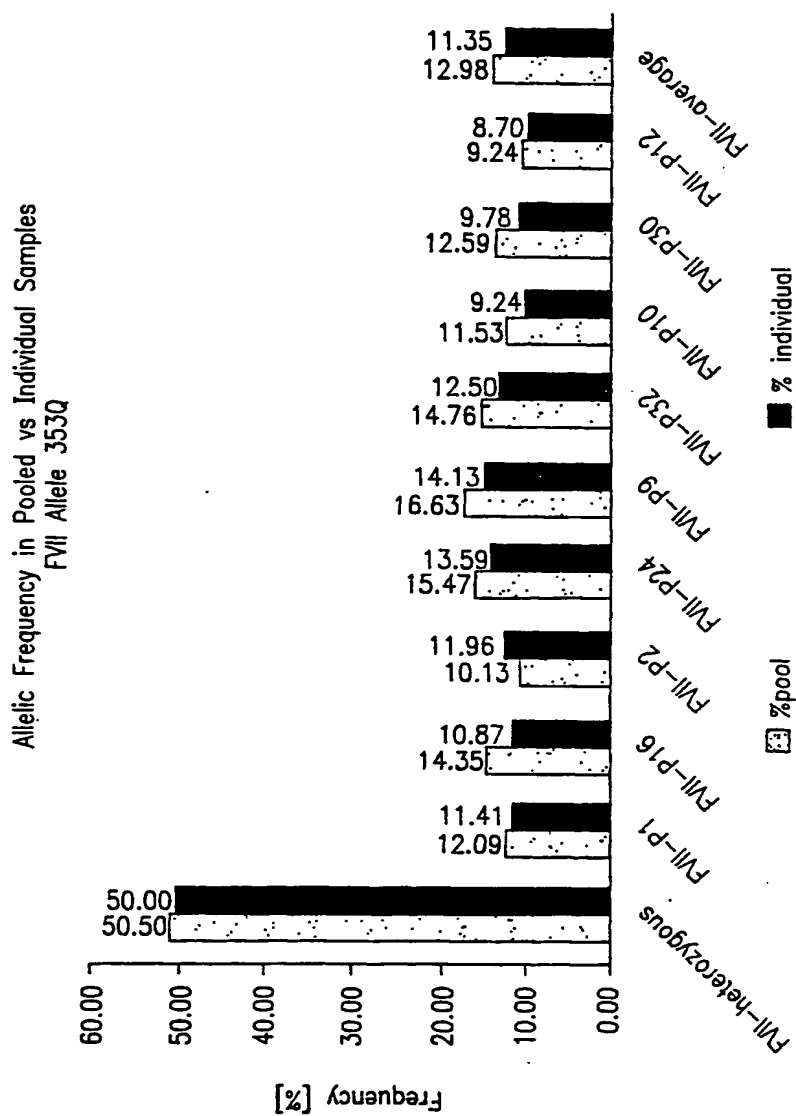


FIG. 9

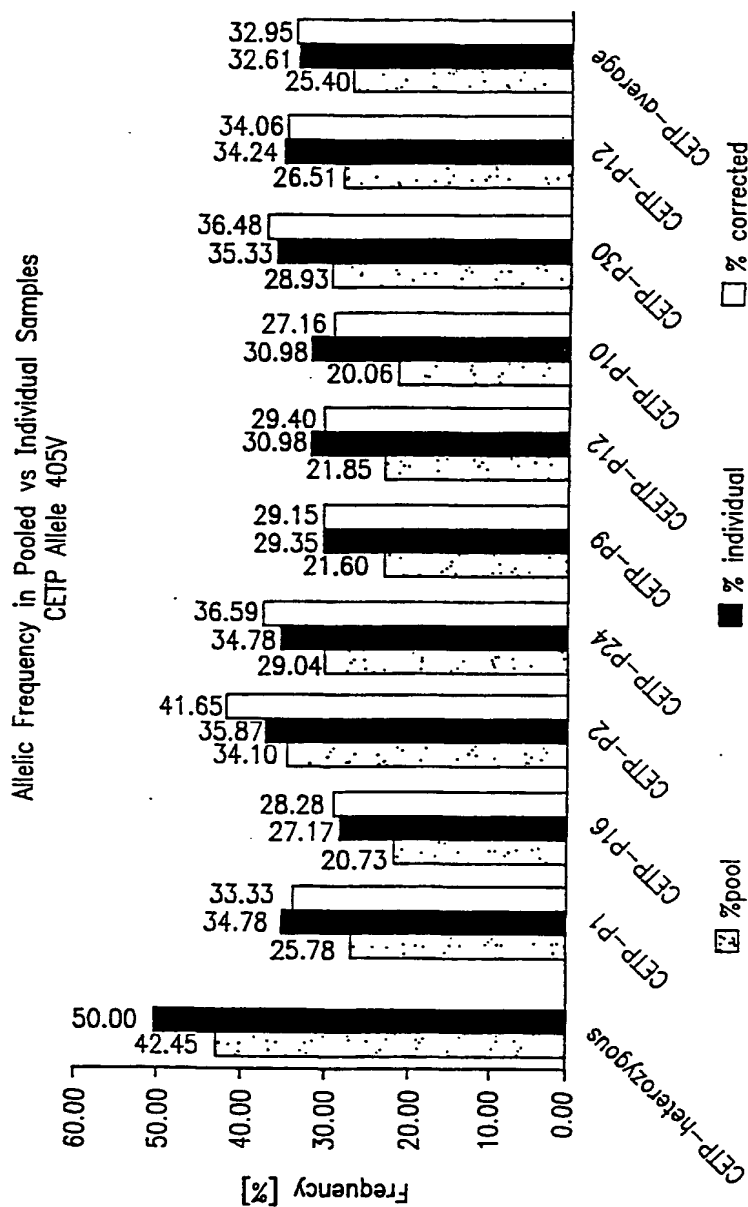


FIG. 10

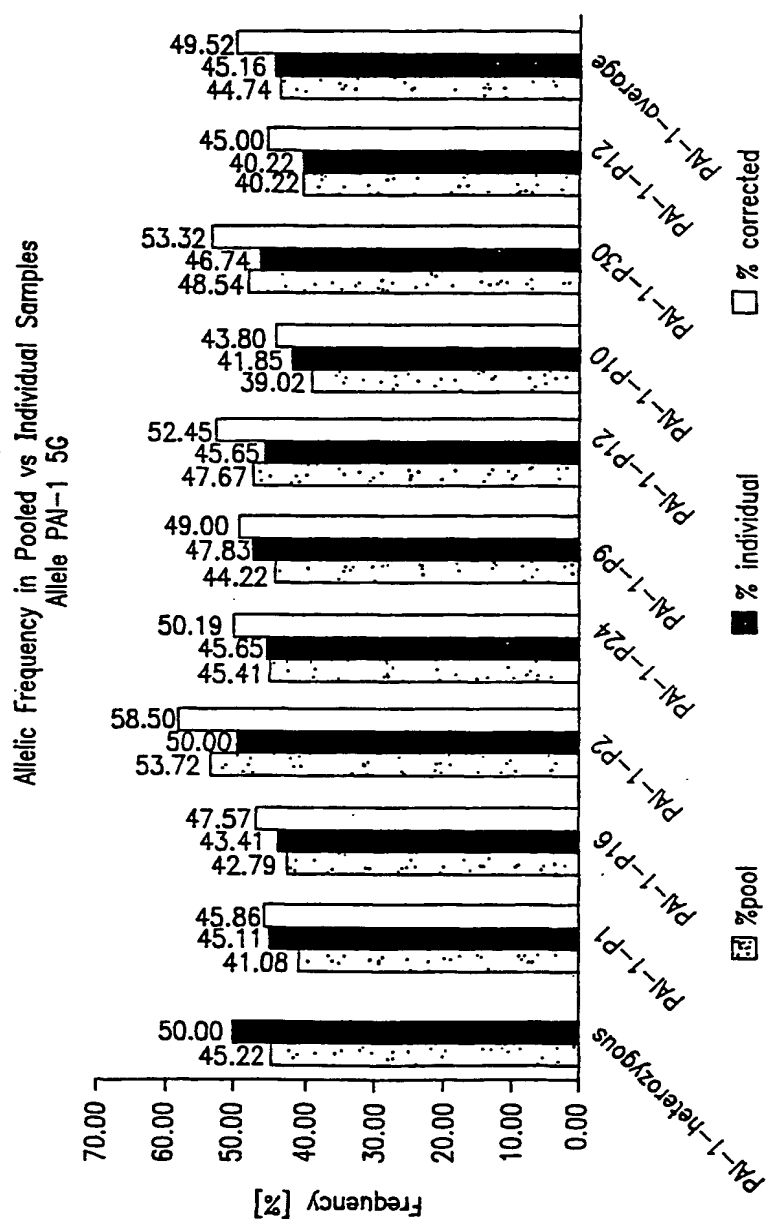


FIG. 11

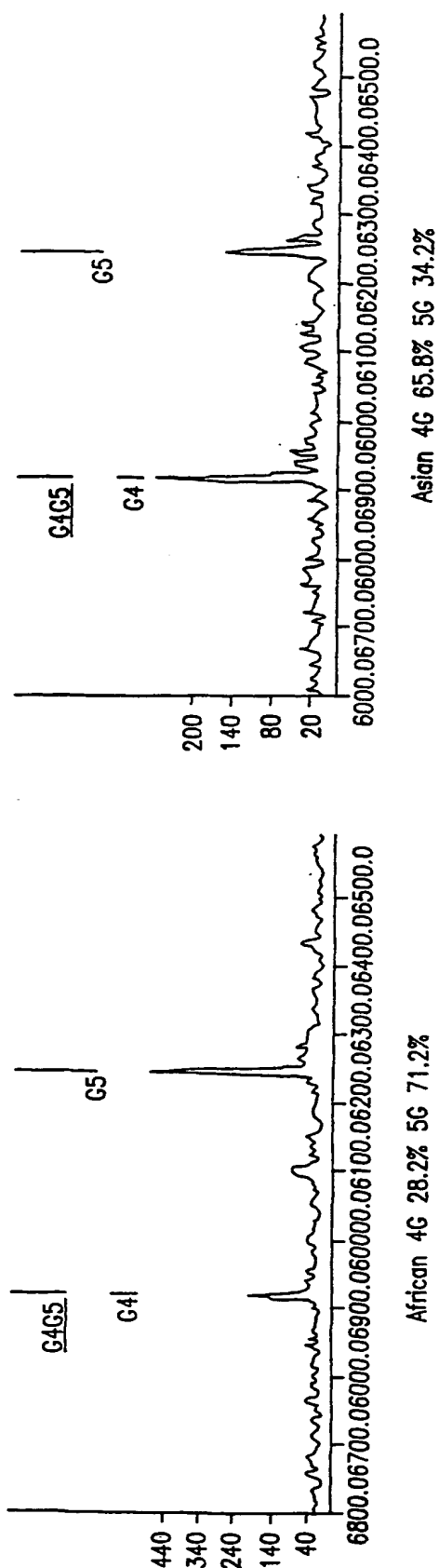


FIG. 12A

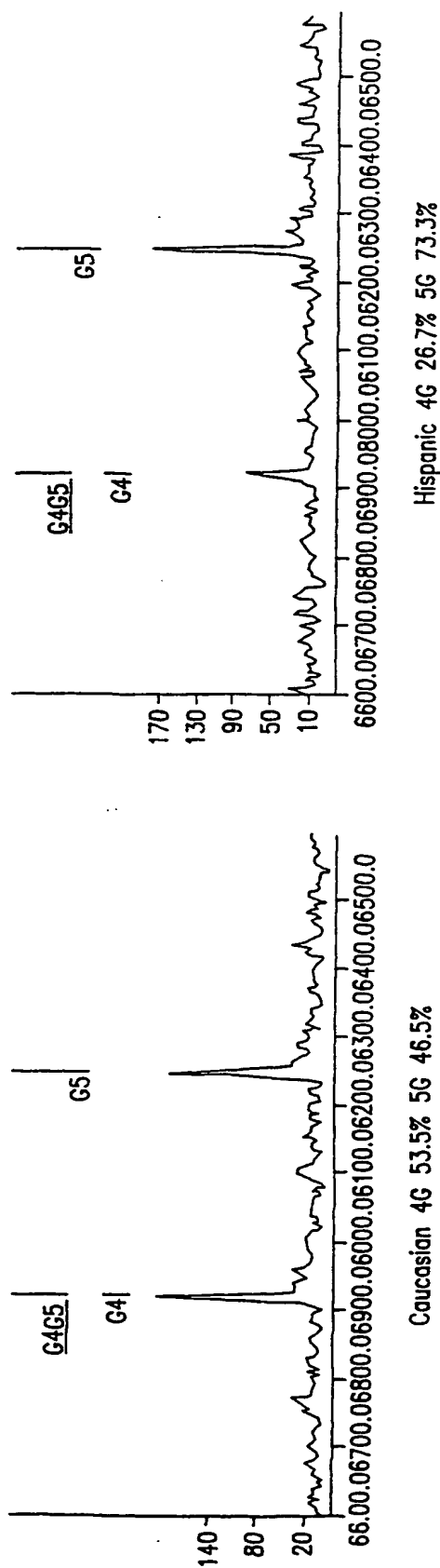


FIG. 12C

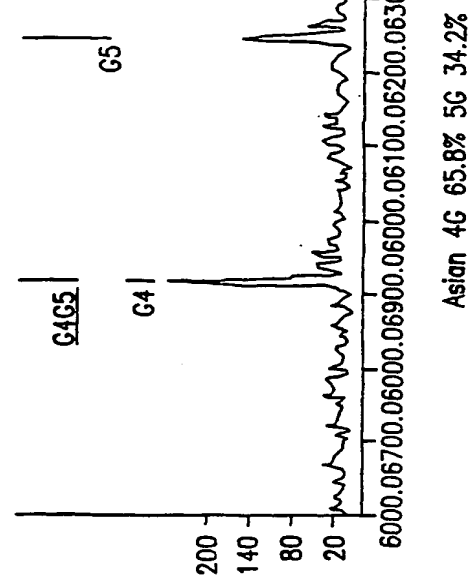


FIG. 12B

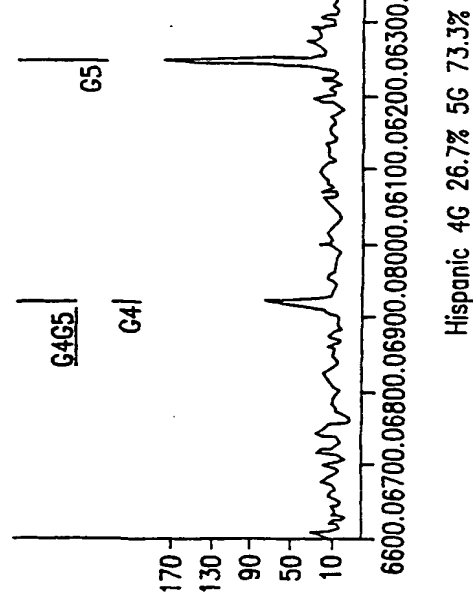
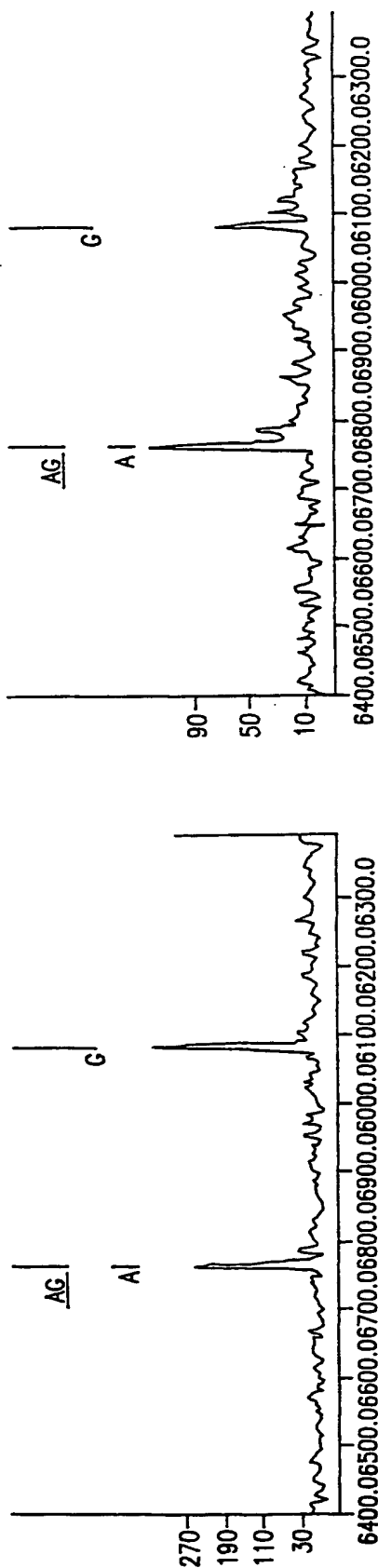


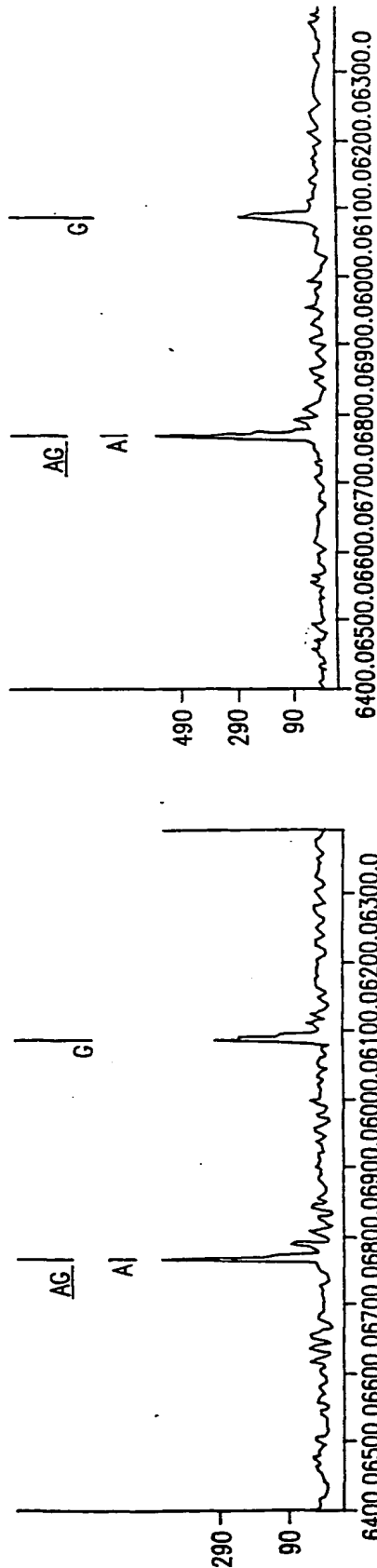
FIG. 12D



African Ile 43.6% Val 56.4%

FIG. 13A

FIG. 13B



Caucasian Ile 63.0% Val 37.0%

FIG. 13C

Hispanic Ile 70.1% Val 29.9%

FIG. 13D

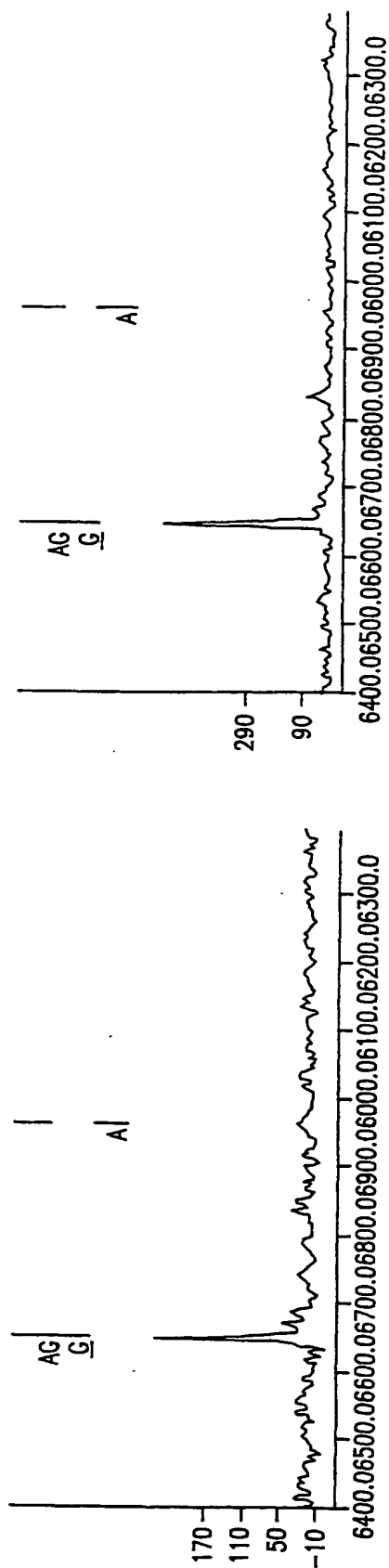


FIG. 14A

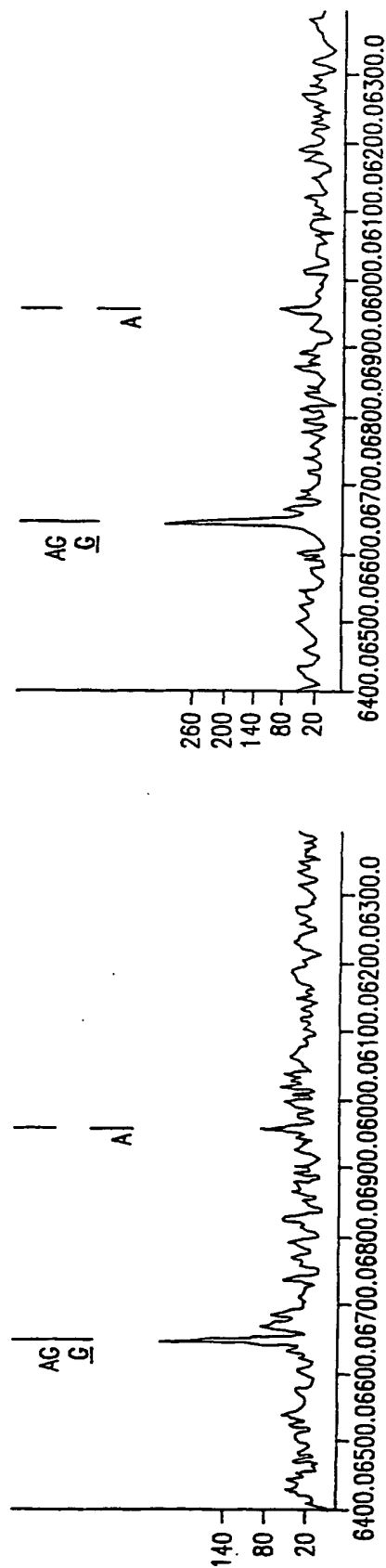


FIG. 14B

Caucasian Arg 88.5% Glu 11.5%

FIG. 14C

Hispanic Arg 89.2% Glu 10.8%

FIG. 14D

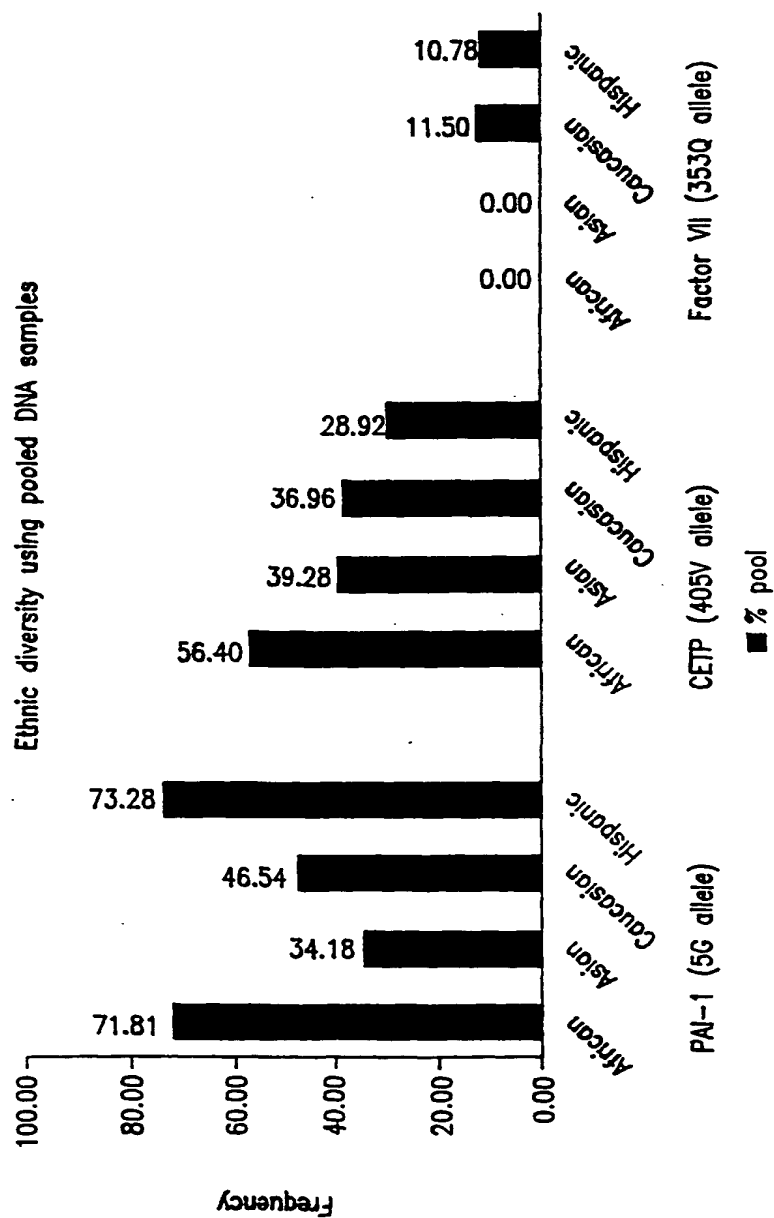


FIG. 15

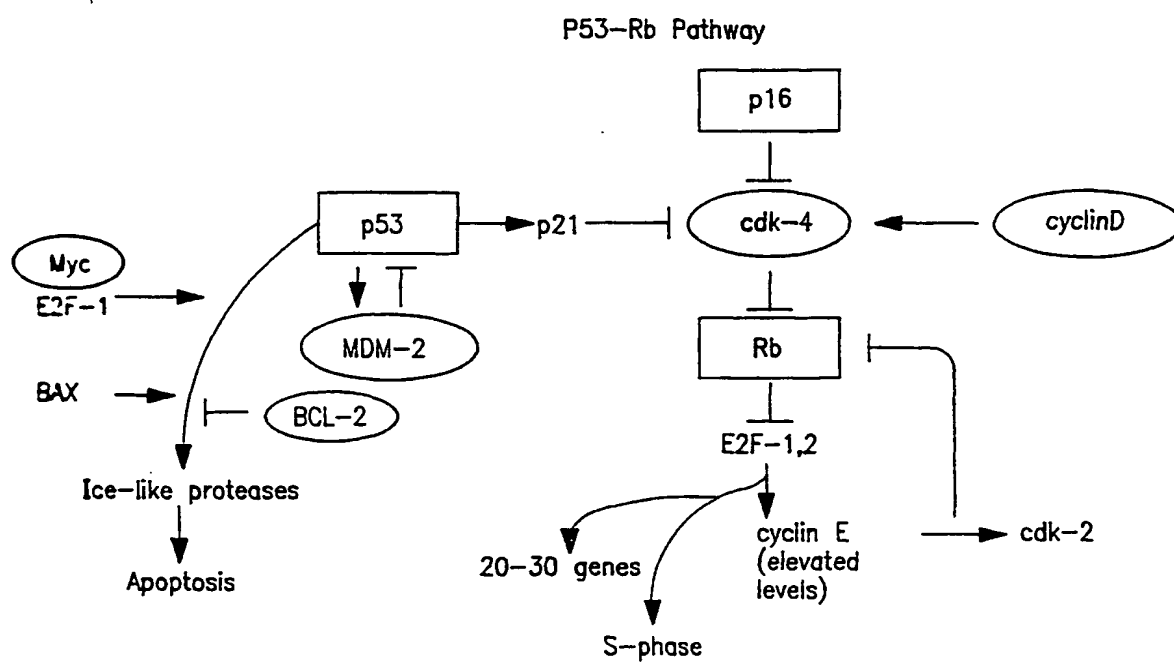


FIG. 16

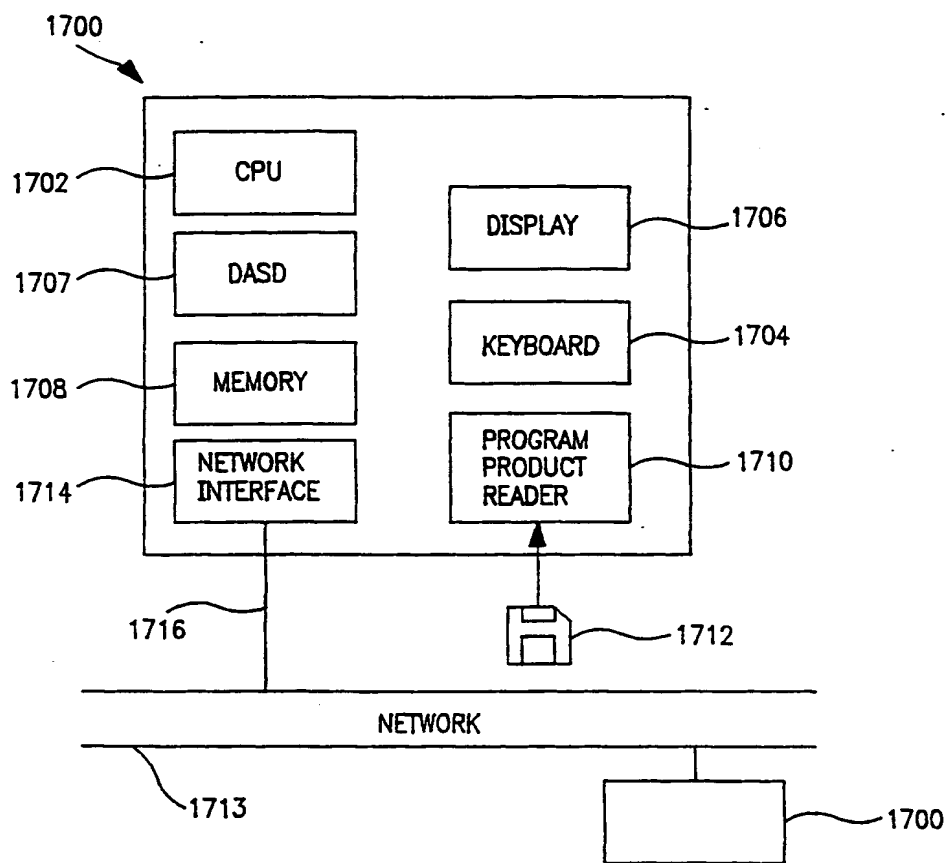


FIG. 17

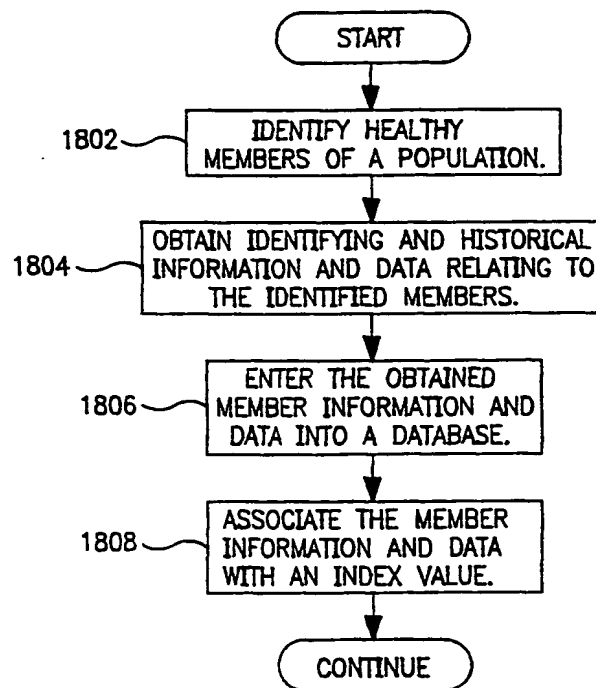


FIG. 18

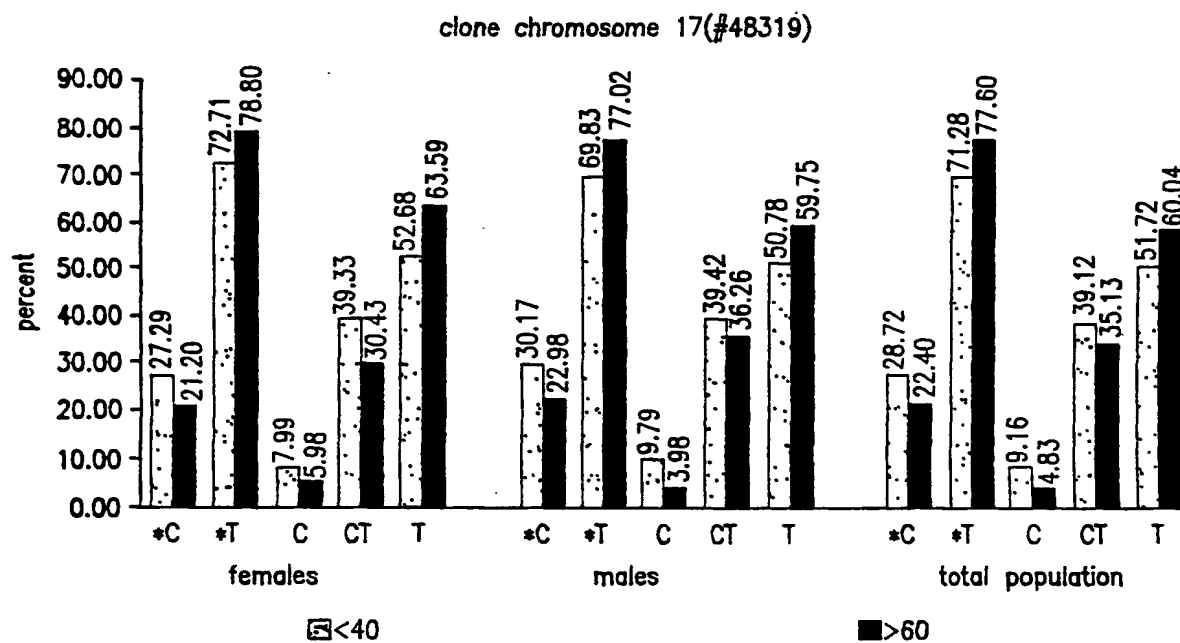


FIG. 19

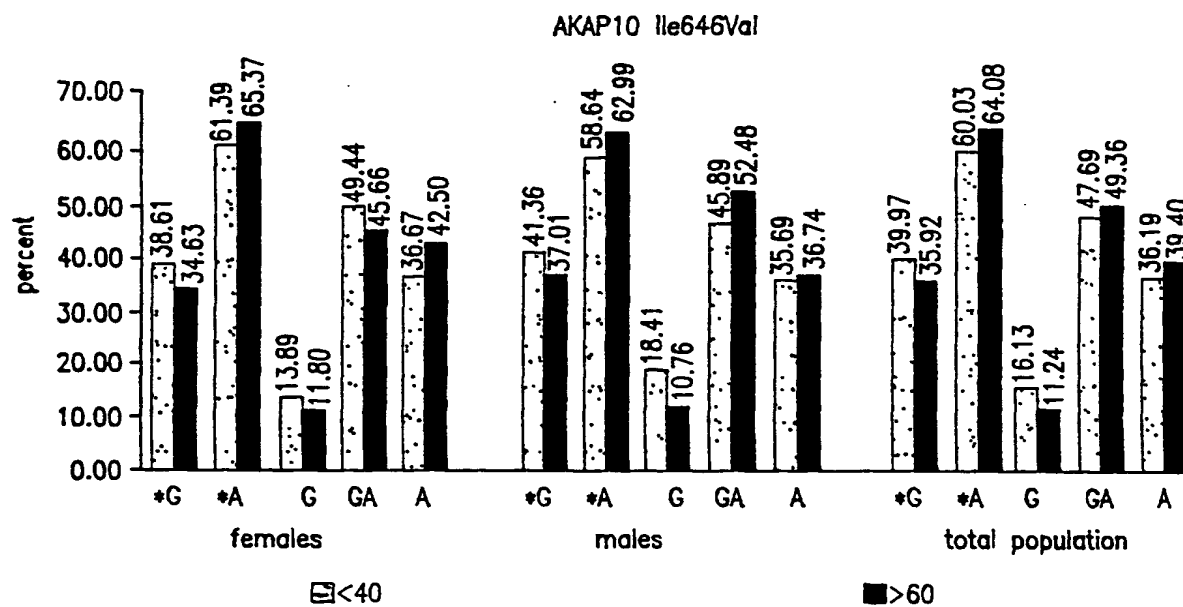


FIG. 20

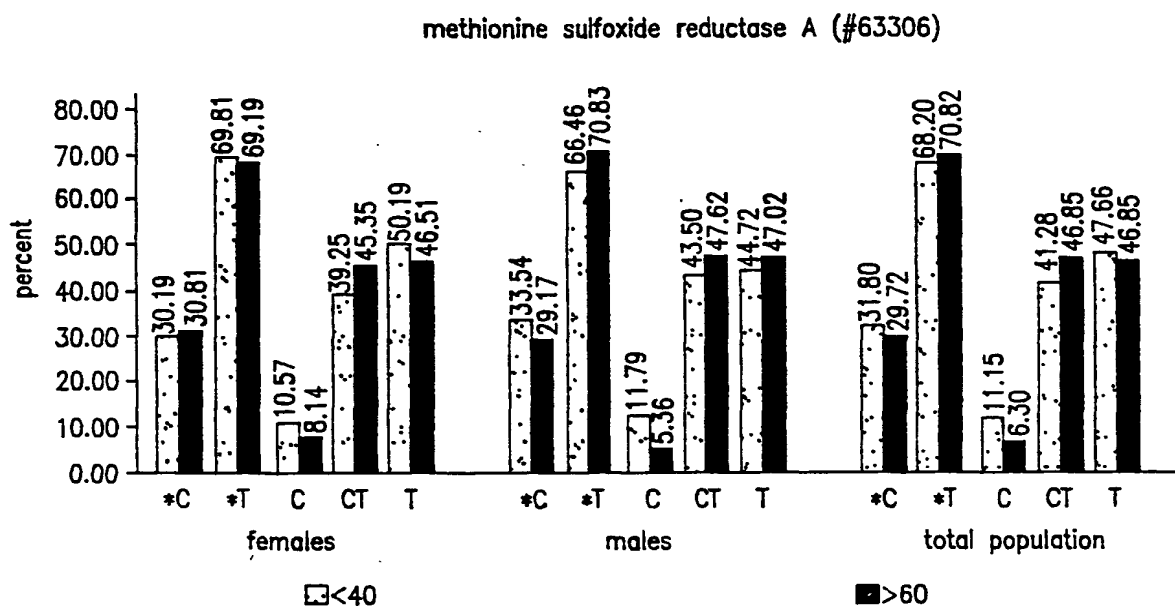


FIG. 2I

Collection Information									
Consent Form Signed <input type="checkbox"/> Yes <input type="checkbox"/> No			Time of Sample Collection (nearest hour, in 24 hour clock format)		Initials		Initials of Data Collector		
Date of Collection							(DO NOT COMPLETE; for data entry only)		
Month	Day	Year					Sample;	<input type="checkbox"/> Intact	Volume (ml)
JAN							<input type="checkbox"/> Lost		
FEB							<input type="checkbox"/> Broken		
MAR									
APR									
MAY									
JUN									
JUL									
AUG									
SEP									
OCT									
NOV									
DEC									

Donor Information									
Date of Birth			Height		Weight		What Physical activity do you do on a regular basis?		Are you a vegetarian?
Month	Day	Year	Feet	Inches	Pounds			<input type="checkbox"/> Running	<input type="checkbox"/> Yes
JAN								<input type="checkbox"/> Swimming	<input type="checkbox"/> No
FEB								<input type="checkbox"/> Biking	
MAR								<input type="checkbox"/> Gymnastics	
APR								<input type="checkbox"/> Other	
MAY								<input type="checkbox"/> None	
JUN									
JUL									
AUG									
SEP									
OCT									
NOV									
DEC									

To the best of your knowledge, what is the Ethnic Origin of your:									
<b>Father</b>		<b>Mother</b>							
<input type="checkbox"/>	<input type="checkbox"/>	Caucasian (please mark specific geographic area below if known)							
<input type="checkbox"/>	<input type="checkbox"/>	Northern Europe (Austria, Denmark, Finland, France, Germany, Netherlands, Norway, Sweden, Switzerland, UK)							
<input type="checkbox"/>	<input type="checkbox"/>	Southern Europe (Greece, Italy, Spain, Turkey)							
<input type="checkbox"/>	<input type="checkbox"/>	Eastern Europe (Czechoslovakia, Hungary, Poland, Russia, Yugoslavia)							
<input type="checkbox"/>	<input type="checkbox"/>	Middle Eastern (Israel, Egypt, Iran, Iraq, Jordan, Syria, Other Arab States)							
<input type="checkbox"/>	<input type="checkbox"/>	African-American							
<input type="checkbox"/>	<input type="checkbox"/>	Hispanic (please mark specific geographic area below if known)							
<input type="checkbox"/>	<input type="checkbox"/>	Mexico							
<input type="checkbox"/>	<input type="checkbox"/>	Central America, South America							
<input type="checkbox"/>	<input type="checkbox"/>	Cuba, Puerto Rico, other Caribbean							
<input type="checkbox"/>	<input type="checkbox"/>	Asian (please mark specific geographic area below if known)							
<input type="checkbox"/>	<input type="checkbox"/>	Japanese							
<input type="checkbox"/>	<input type="checkbox"/>	Chinese							
<input type="checkbox"/>	<input type="checkbox"/>	Korean							
<input type="checkbox"/>	<input type="checkbox"/>	Vietnamese							
<input type="checkbox"/>	<input type="checkbox"/>	Filipino							
<input type="checkbox"/>	<input type="checkbox"/>	Native American							
<input type="checkbox"/>	<input type="checkbox"/>	Other _____							
<input type="checkbox"/>	<input type="checkbox"/>	Don't know							

How long have you lived there?									
Years	What is your highest grade you completed in school?								
	less than 8th grade								
	8th, 9th, 10th, or 11th grade								
	high school graduate or equivalency								
	some college, 2yr degree								
	college graduate, 4yr degree								
	post graduate education or degree								

Mother Deceased? Cause of Death Mother: Father Deceased? Cause of Death Father:									
<input type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> Yes							
<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No							
If Yes at what age?		< 29	Heart Disease		If Yes at what age?		< 29	Heart Disease	
		30-39	Cancer				30-39	Cancer	
		40-49	Stroke				40-49	Stroke	
		50-59	Accident				50-59	Accident	
		60-69	Suicide				60-69	Suicide	
		70-79	Other, _____				70-79	Other, _____	
		80-89					80-89		
		≥ 90					> 90		

**EIG 22A**

FIG. 22A

**SUBSTITUTE SHEET (RULE 26)**

Have you ever smoked? ☐ Yes ☐ No

If yes, for how long?

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

Years

Have you been hospitalized in the past 5 years for more than 6 days at a time?  
☐ Yes ☐ No

If yes, how many times?

0	1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---	---

For each hospitalization (if not the same) how long did you stay and for what reason?

1) Weeks: 

0	1	2	3	4	5	6
---	---	---	---	---	---	---

☐ Acute disorder, including infection and thrombosis☐ Chronic Disorder☐ Accident☐ Other: \_\_\_\_\_2) Weeks: 

0	1	2	3	4	5	6
---	---	---	---	---	---	---

☐ Acute disorder, including infection and thrombosis☐ Chronic Disorder☐ Accident☐ Other: \_\_\_\_\_3) Weeks: 

0	1	2	3	4	5	6
---	---	---	---	---	---	---

☐ Acute disorder, including infection and thrombosis☐ Chronic Disorder☐ Accident☐ Other: \_\_\_\_\_

Have you or has anyone in your immediate family (parents, brothers, sisters, or your children) had the following?

Mark all that apply!

Disease

Heart Disease, including arteriosclerosis
Stroke
Hypertension
Blood clots
Diabetes, insulin dependent
Diabetes, not insulin-dependent (diet controlled)
Cancer:
Lung & Bronchus
Breasts
Prostate
Colon & Rectum
Skin
Lymphoma & Leukemia
Other, please specify below:
Alzheimer's Disease
Epilepsy
Schizophrenia
Bipolar disorder (manic depression)
Major depression
Chronic Inflammatory or Autoimmune Disease including Multiple Sclerosis and Rheumatoid Arthritis
Emphysema
Asthma
Other, please specify below:

You    Mother    Father    Sister    Brother    Child

☐ ☐ ☐ ☐ ☐ ☐

Do you take prescription drugs on a regular basis?

☐ Yes ☐ No

If yes, please specify below:

Have you ever donated blood before? ☐ Yes ☐ No

If yes, how many times:    Number of Times

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

Additional health information details you would like to provide:


FOR  
OFFICE  
USE ONLY

0	1
2	3
4	5
6	7
8	9

Do you drink any kind of alcoholic beverage?

☐ Never☐ Less than 3 times per week☐ Daily☐ Hardly ever☐ 3 or more times per week

FIG. 22B

SUBSTITUTE SHEET (RULE 26)

DOCID: <WQ 0127957A? 1A>

What is your highest grade you completed in school?

☐ less than 8th grade  
☐ 8th, 9th, 10th, or 11th grade  
☐ high school graduate or equivalency  
☐ some college, 2yr degree  
☐ college graduate, 4yr degree  
☐ post graduate education or degree

Mother Deceased? Cause of Death Mother: Father Deceased? Cause of Death Father:

☐ Yes ☐ No ☐ Yes ☐ No

If Yes at what age? ☐ < 29 ☐ 30-39 ☐ 40-49 ☐ 50-59 ☐ 60-69 ☐ 70-79 ☐ 80-89 ☐ ≥ 90

Heart Disease ☐ Yes ☐ No  
 Cancer ☐ Yes ☐ No  
 Stroke ☐ Yes ☐ No  
 Accident ☐ Yes ☐ No  
 Suicide ☐ Yes ☐ No  
 Other, \_\_\_\_\_

## Health Information

Have you or has anyone in your immediate family (parents, brothers, sisters, or your children) had the following? Mark all that apply

Disease	You	Mother	Father	Sister	Brother	Child
Heart Disease	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stroke	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hypertension	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Blood clots	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diabetes, insulin dependent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diabetes, no: insulin-dependent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cancer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lung & Bronchus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Breasts	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Prostate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Colon & Rectum	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stomach	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lymphoma & Leukemia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other, please specify below:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Alzheimer's Disease	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Epilepsy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Schizophrenia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bipolar disorder (manic depression)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Major depression	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Chronic Inflammatory or Autoimmune Disease including Multiple Sclerosis and Rheumatoid Arthritis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Emphysema	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Asthma	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other, please specify below:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Do you take prescription drugs on a regular basis?

If yes, please specify below:

☐ Yes ☐ No Have you ever donated blood before? ☐ Yes ☐ No

If yes, how many times:

Number of Times

Have you been hospitalized in the past 5 years for more than 6 days at a time?  
☐ Yes ☐ No

If yes, how many times?  
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

For each hospitalization (if not the same) how long did you stay and for what reason?

1) Weeks: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

☐ Acute disorder, including infection and thrombosis  
☐ Chronic Disorder  
☐ Accident  
☐ Other: \_\_\_\_\_

2) Weeks: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

☐ Acute disorder, including infection and thrombosis  
☐ Chronic Disorder  
☐ Accident  
☐ Other: \_\_\_\_\_

3) Weeks: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

☐ Acute disorder, including infection and thrombosis  
☐ Chronic Disorder  
☐ Accident  
☐ Other: \_\_\_\_\_

Do you drink any kind of alcoholic beverage?

☐ Never ☐ Hardly ever  
☐ Less than 3 times per week ☐ 3 or more times per week  
☐ Daily

Additional health information details you would like to provide:

FOR OFFICE USE ONLY

<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>

FIG. 22D

SUBSTITUTE SHEET (RULE 26)

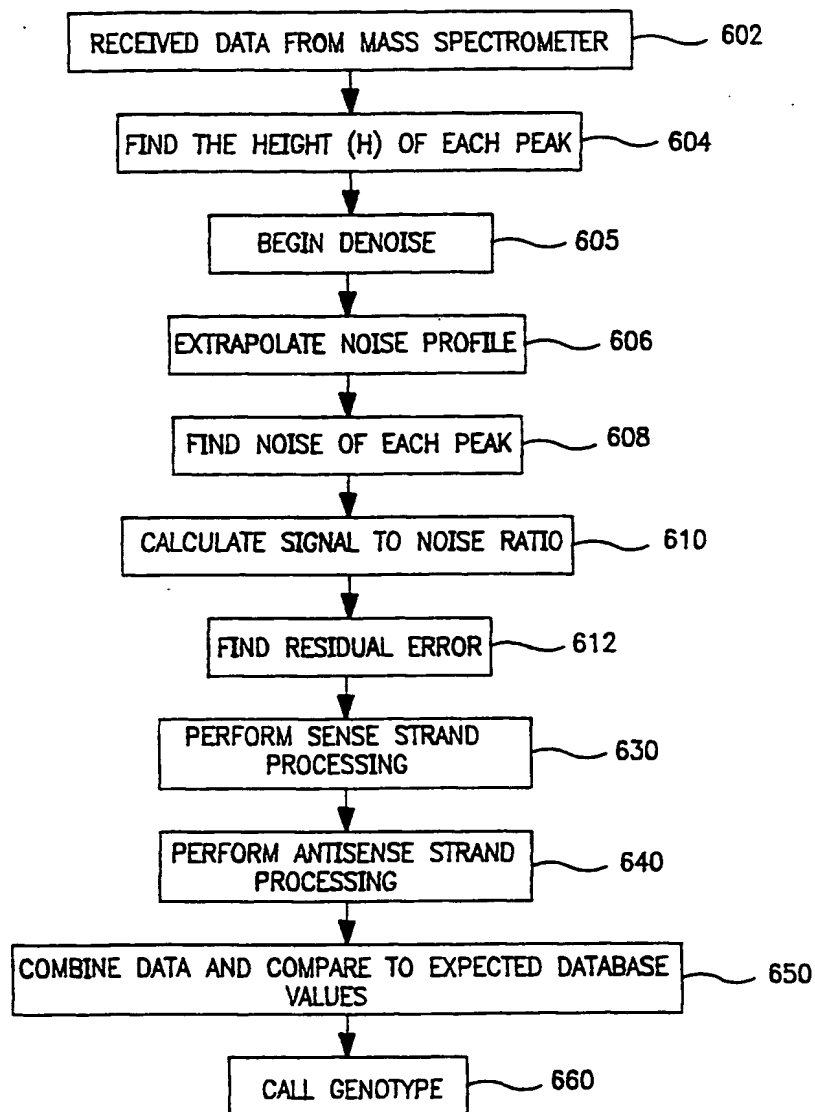


FIG. 23

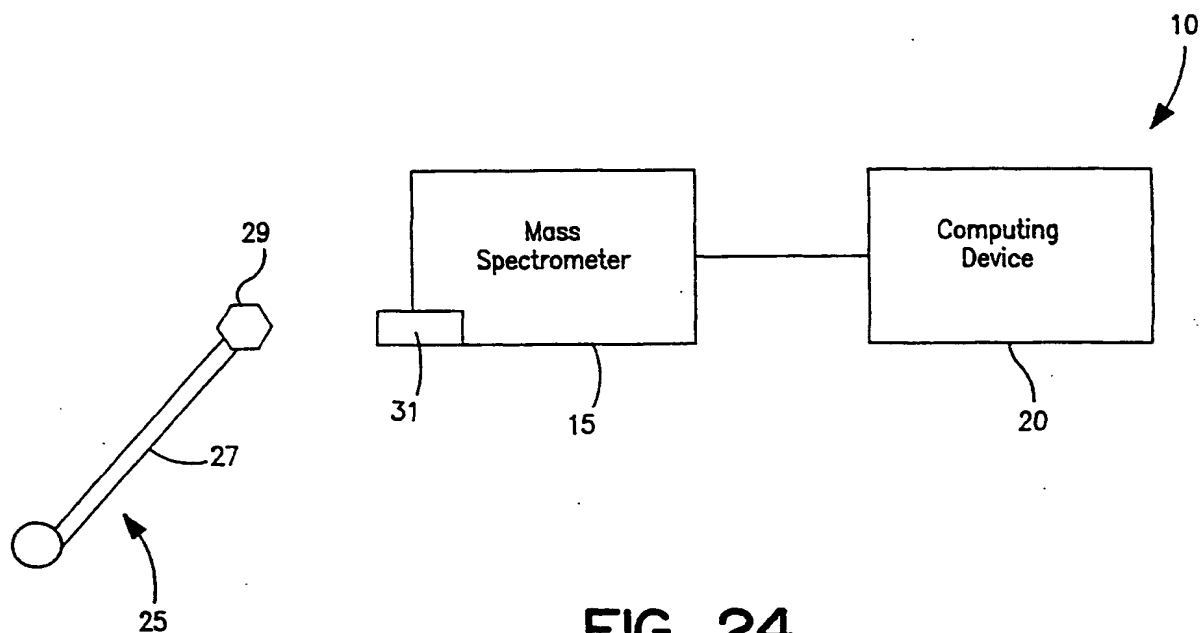
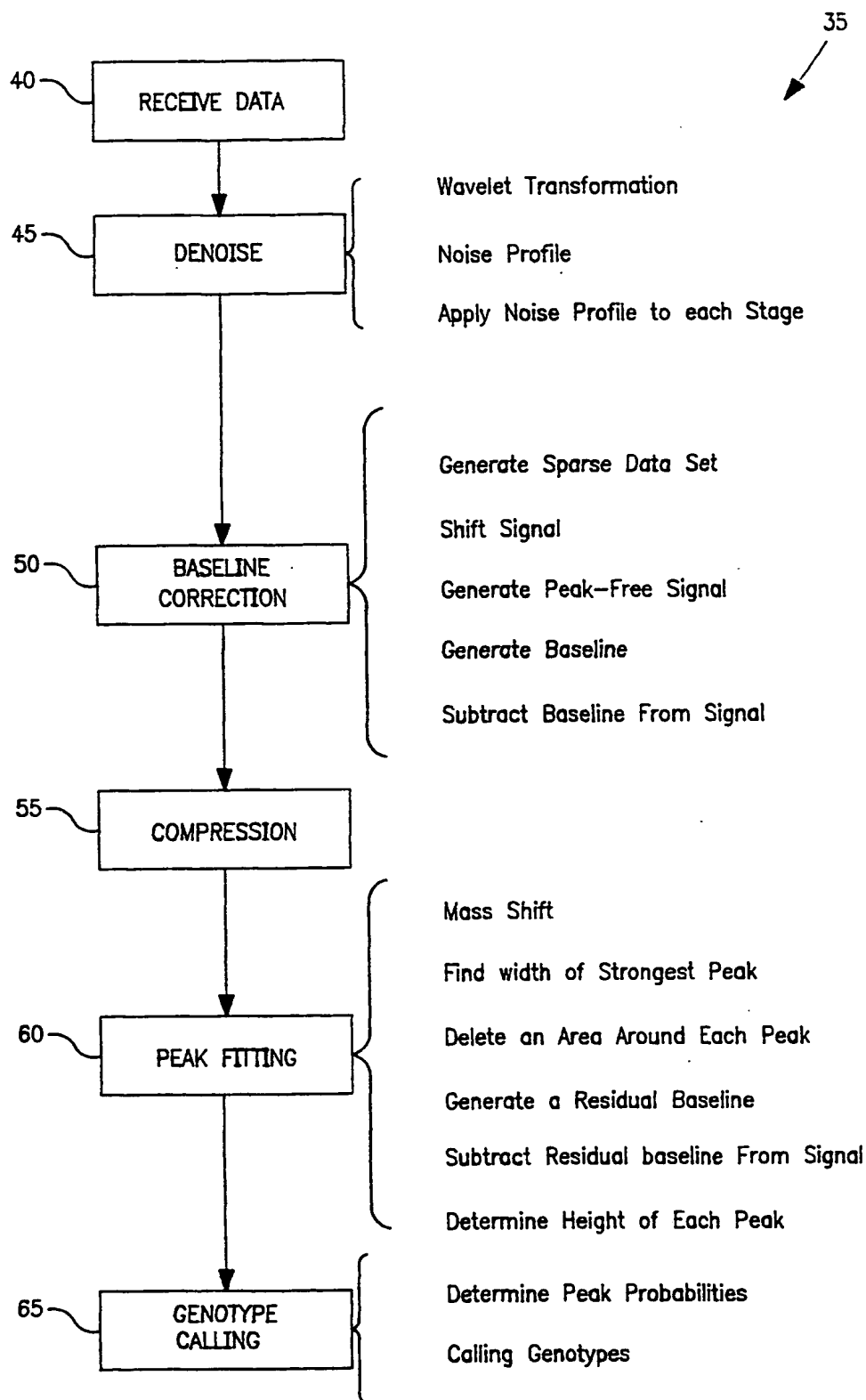


FIG. 24

**FIG. 25**

SUBSTITUTE SHEET (RULE 26)

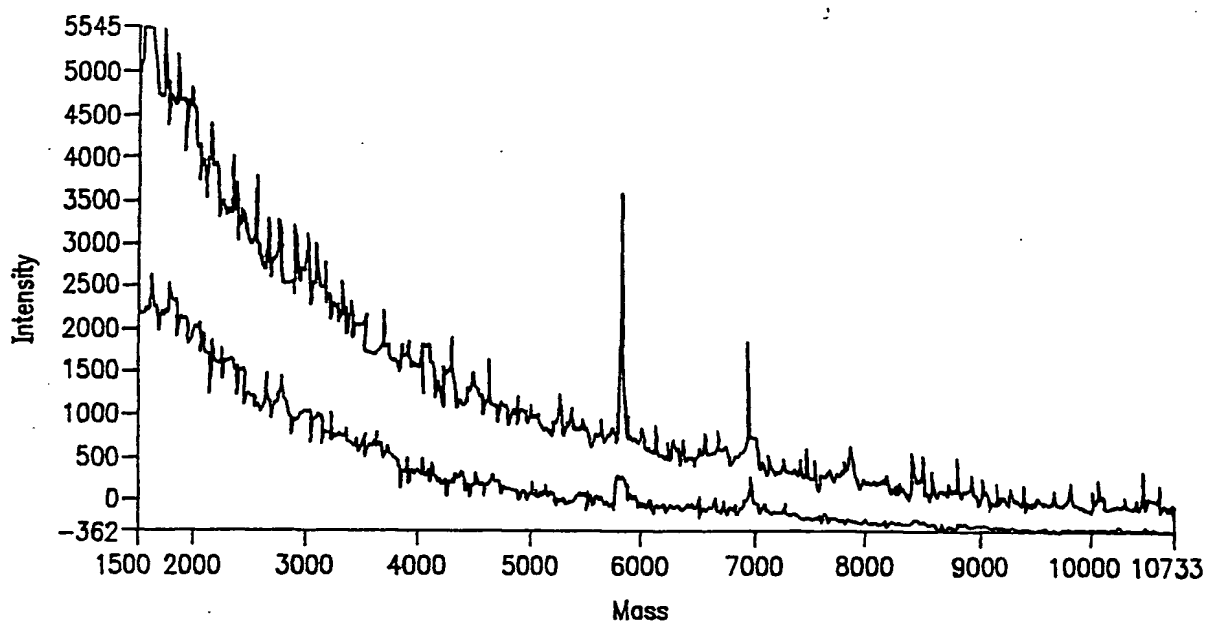


FIG. 26

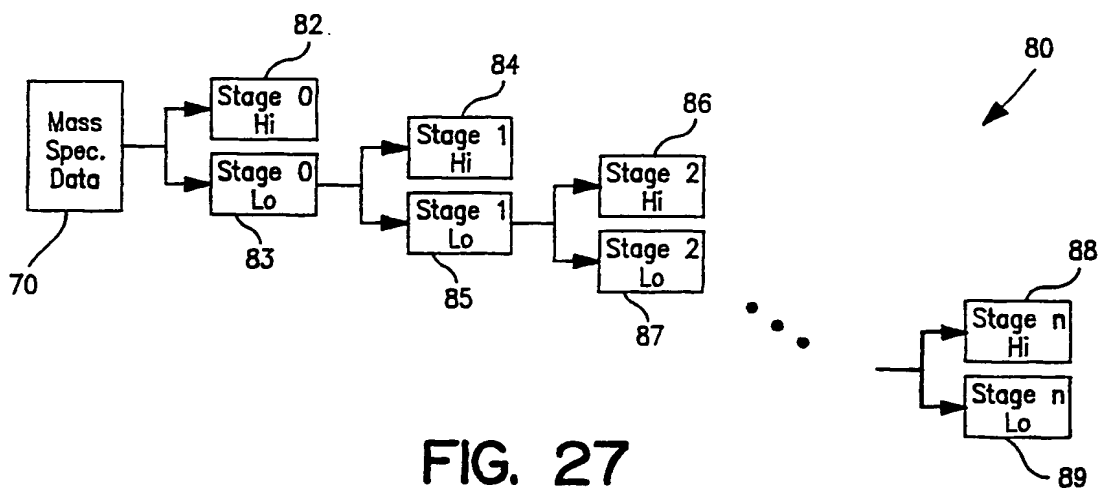


FIG. 27

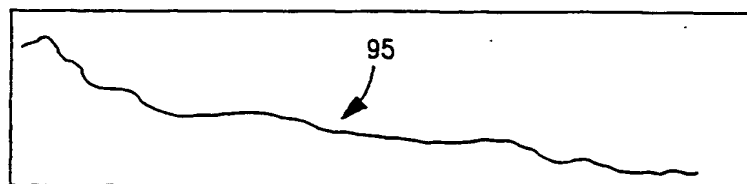


FIG. 28

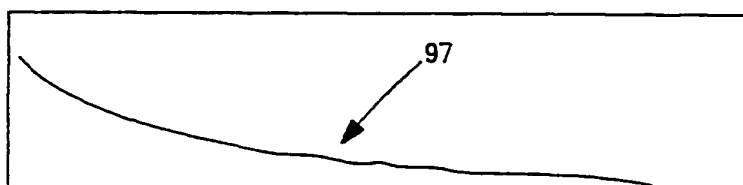


FIG. 29

Exp fitting  
 $a_0 + a_1 \exp^{-a_2 m}$

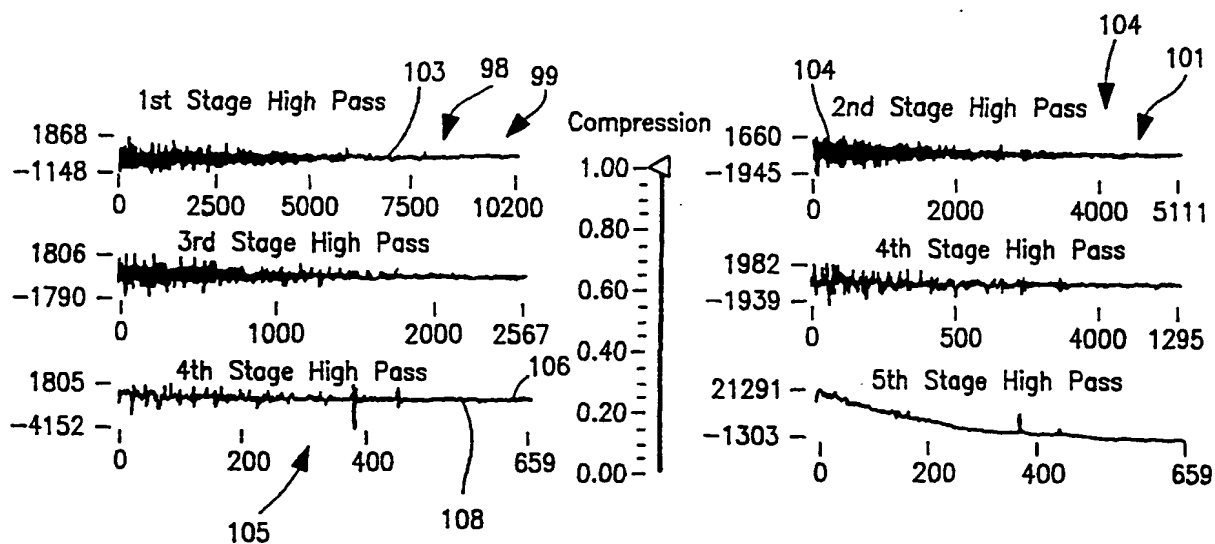


FIG. 30

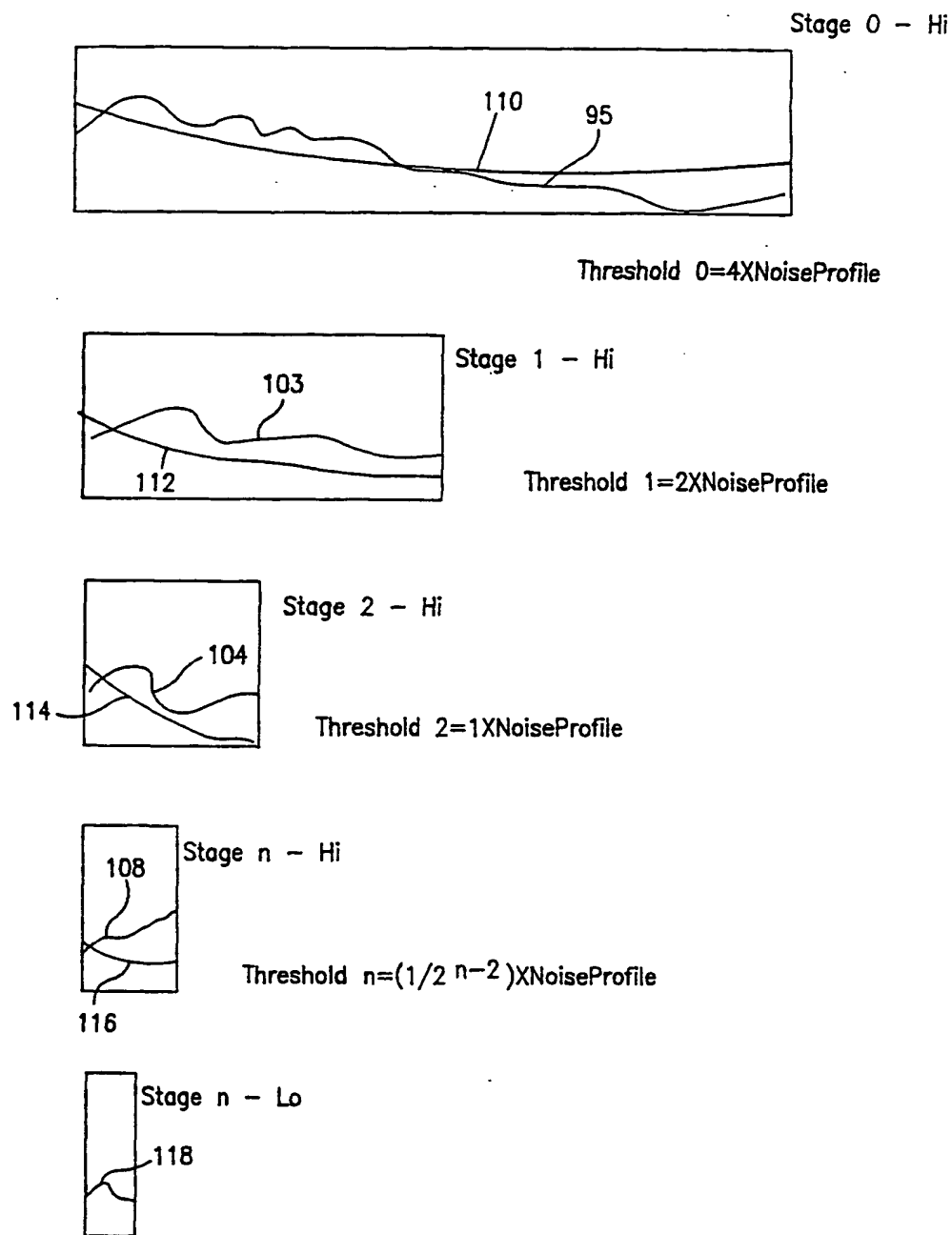


FIG. 3I

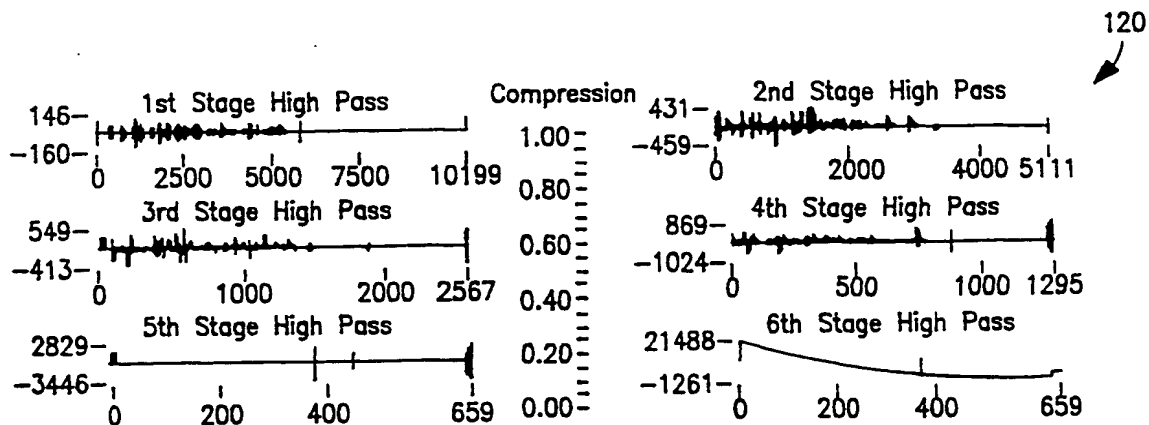


FIG. 32

$$\text{Signal}(t) = \frac{(\text{Start } 0(t) + \text{Start } 1(t) + \text{Start } 2(t) \dots + \text{Start } 23(t))}{24}$$

SHIFT SIGNAL TO ACCOUNT FOR  
VARIATIONS DUE TO STARTING POINT

FIG. 33

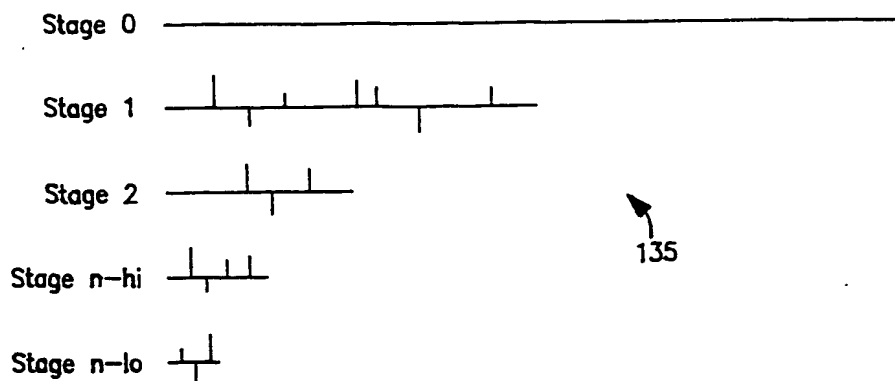


FIG. 34

SUBSTITUTE SHEET (RULE 26)

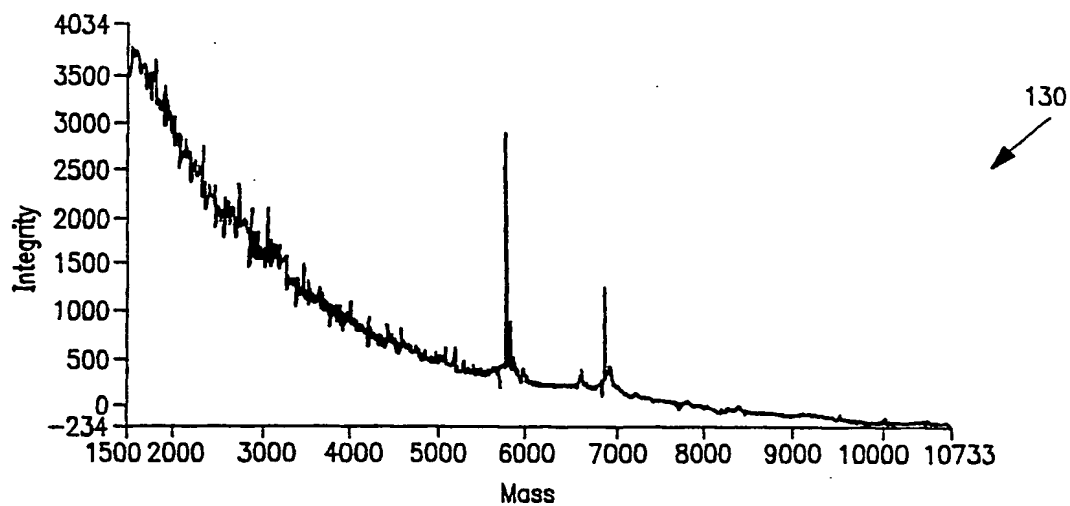


FIG. 35

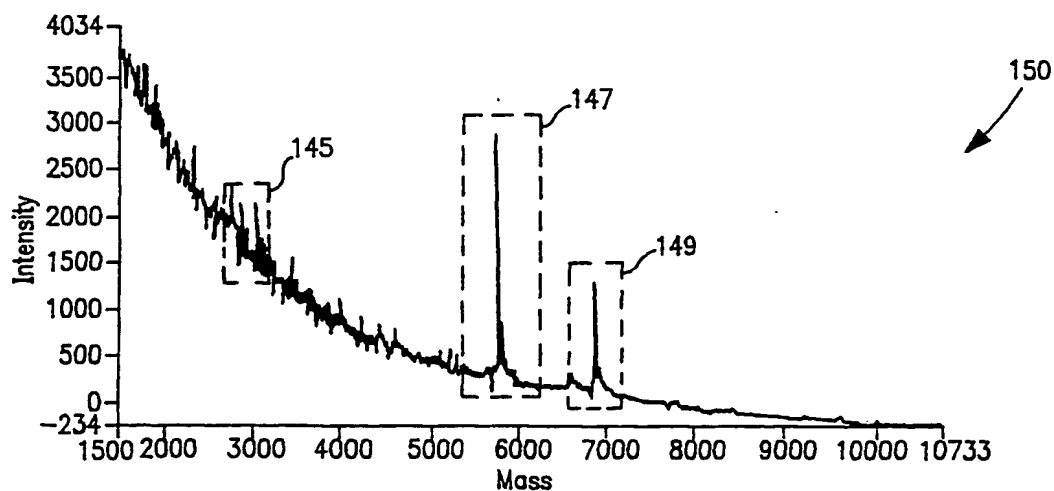


FIG. 13—TAKE A MOVING AVERAGE, REMOVE SECTIONS EXCEEDING A THRESHOLD

FIG. 36

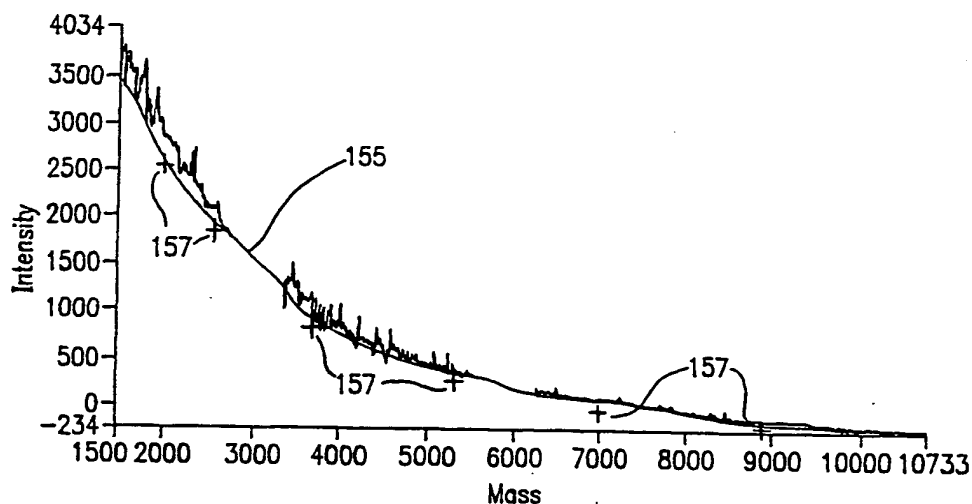


FIG. 37

FIND MINIMA IN REMAINING SIGNALS AND CONNECT TO FORM A PEAK FREE SIGNAL

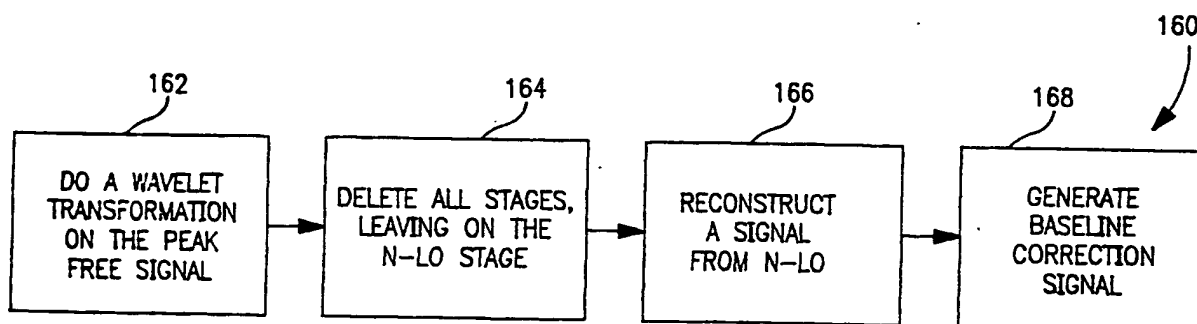


FIG. 38

GENRATE BASELINE CORRECTION

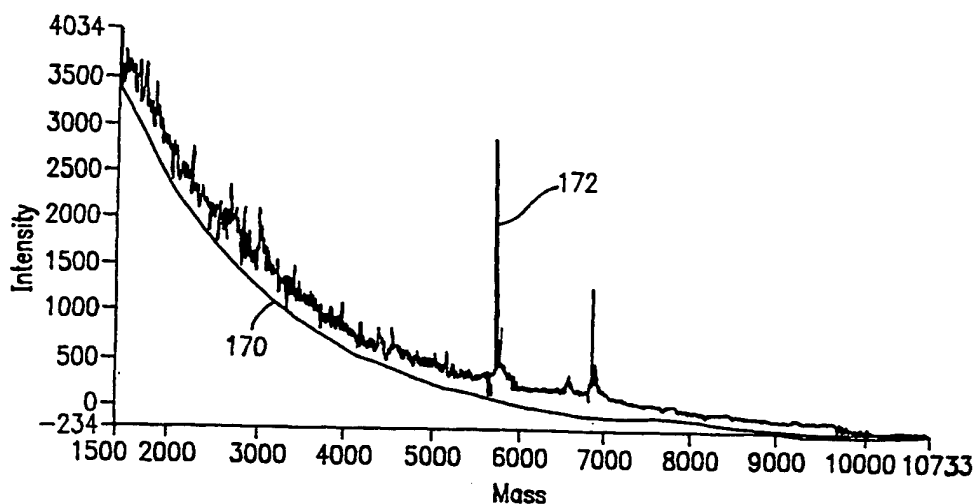


FIG. 39

SUBSTITUTE SHEET (RULE 26)

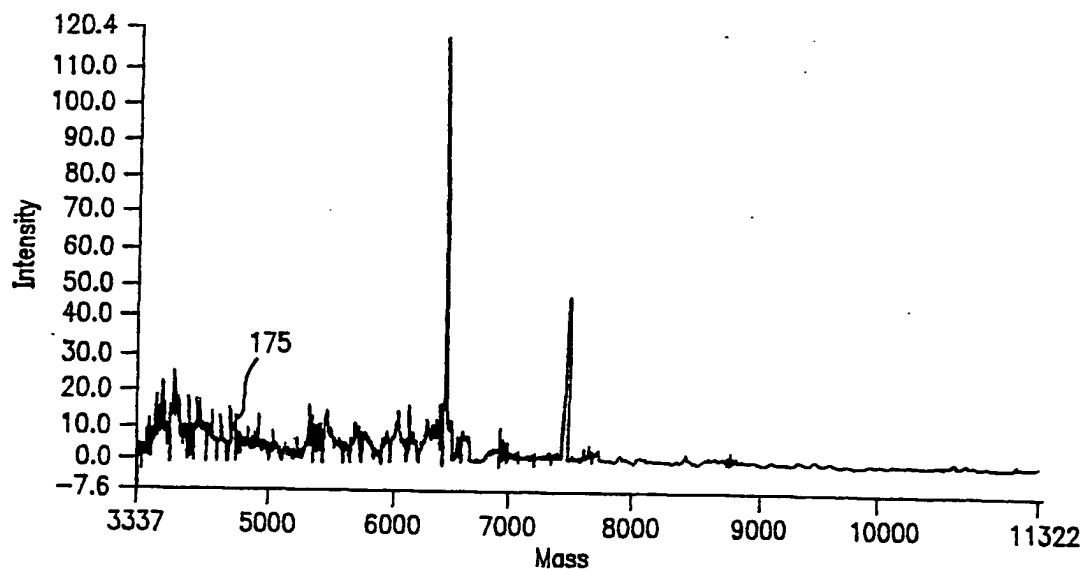


FIG. 40

NON-O COEFFICIENTS	COEFFICIENTS	INTERMEDIATE	RELATIVE
<div>183</div> 100 150 500 10,050 10,075 11,125 12,100 13,250	<div>182</div> <div>184</div> <div>188</div> 25 220 .1 800 890 910 1000 (MAX) 940	<div>186</div> <div>190</div> 100.025 150.220 500.0001 10,050.8 10,075.89 11,125.91 12,100.99999 13,250.94	<div>180</div> <div>195</div> 100.025 50.220 350.0001 9550.8 25.89 150.91 975.99999 1150.94

FIG. 41

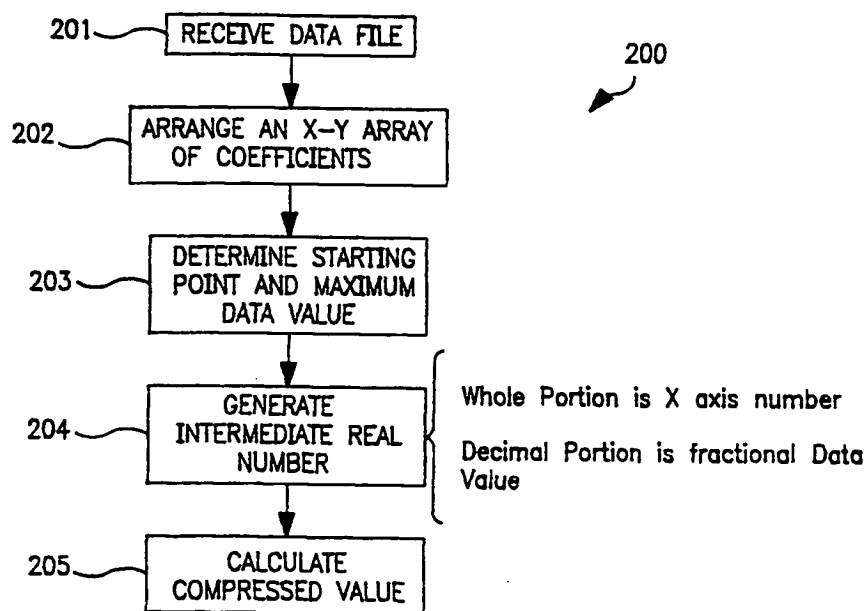


FIG. 42

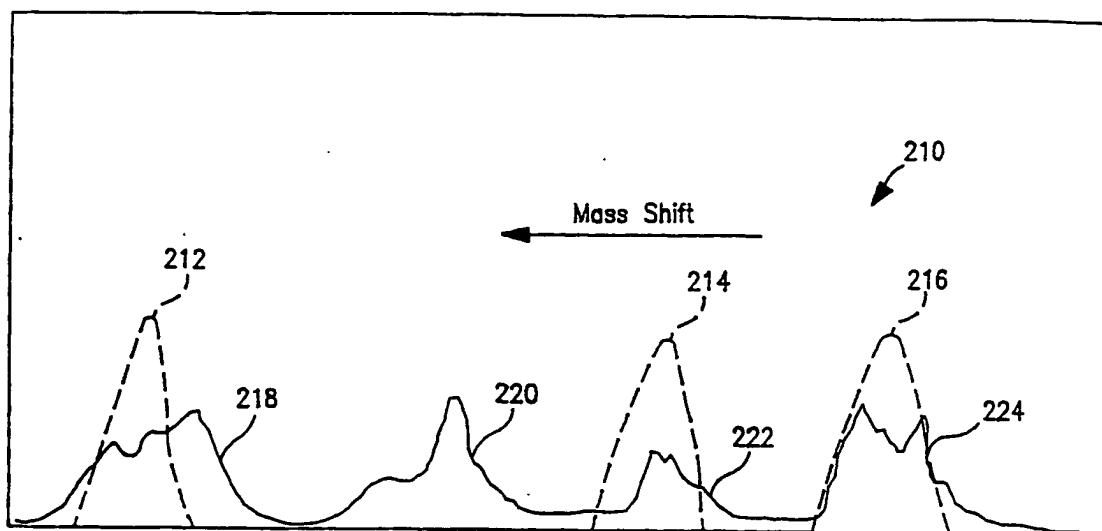


FIG. 43

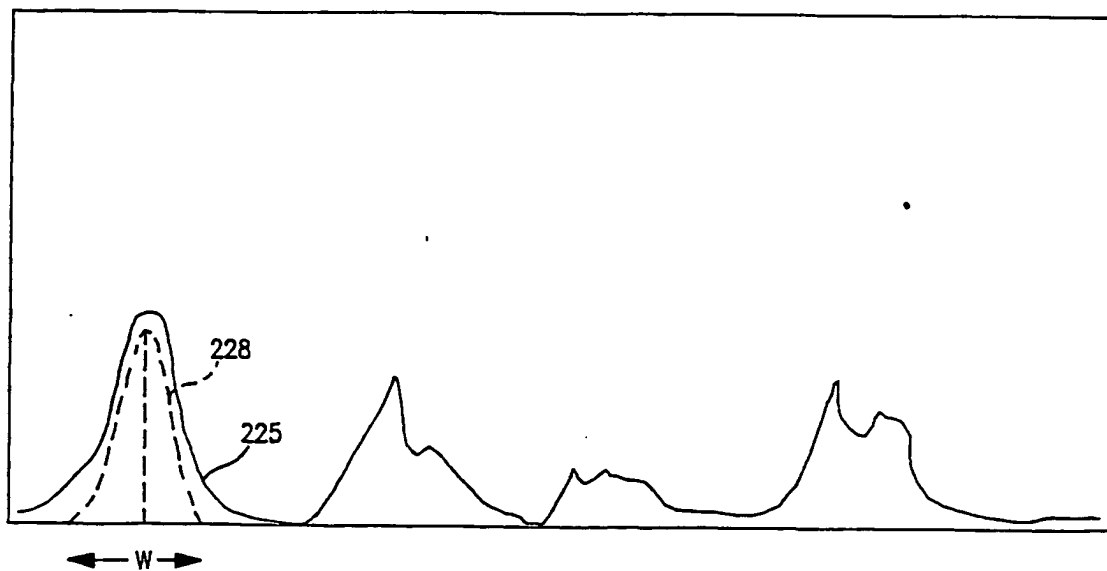


FIG. 44

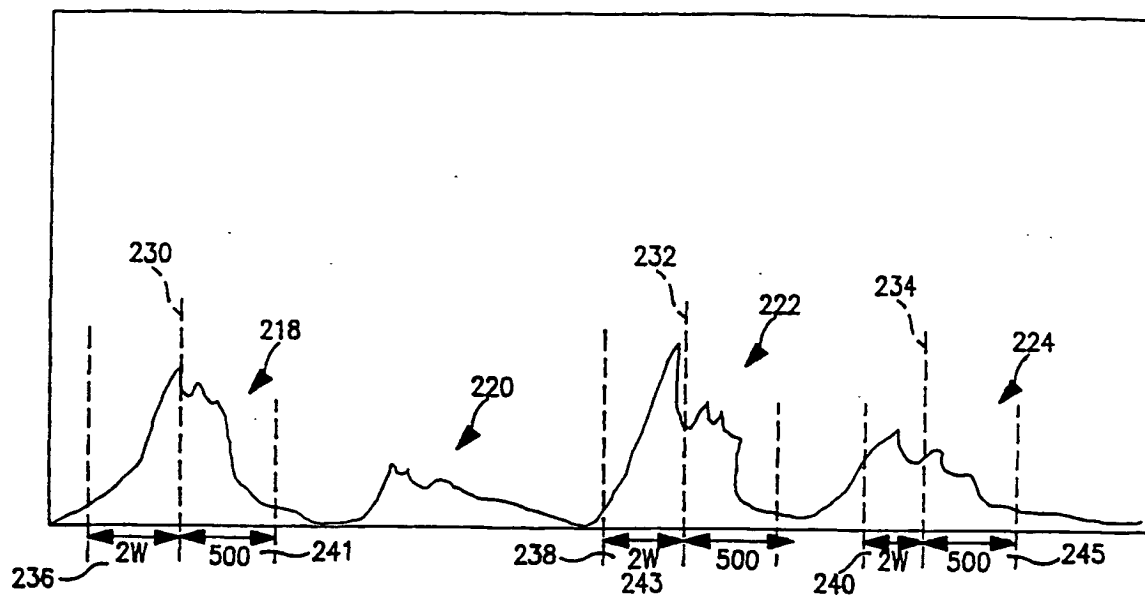


FIG. 45

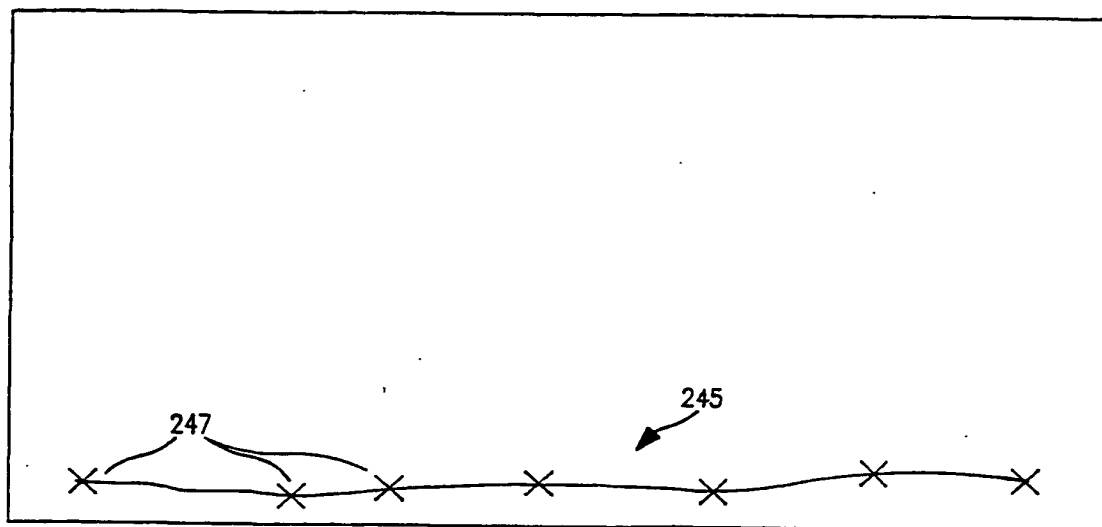


FIG. 46

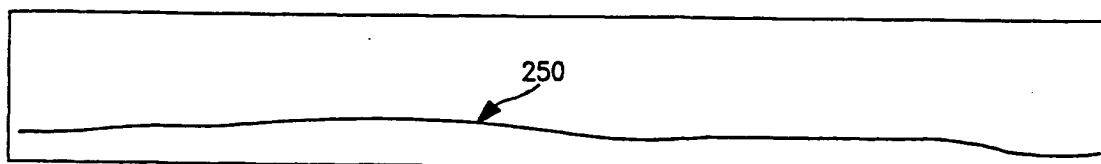


FIG. 47

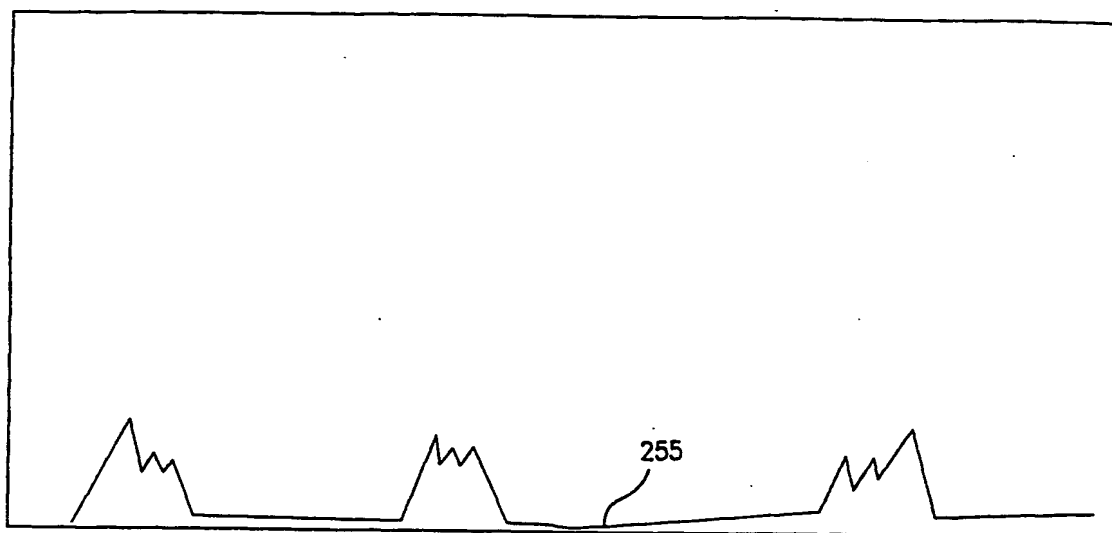


FIG. 48

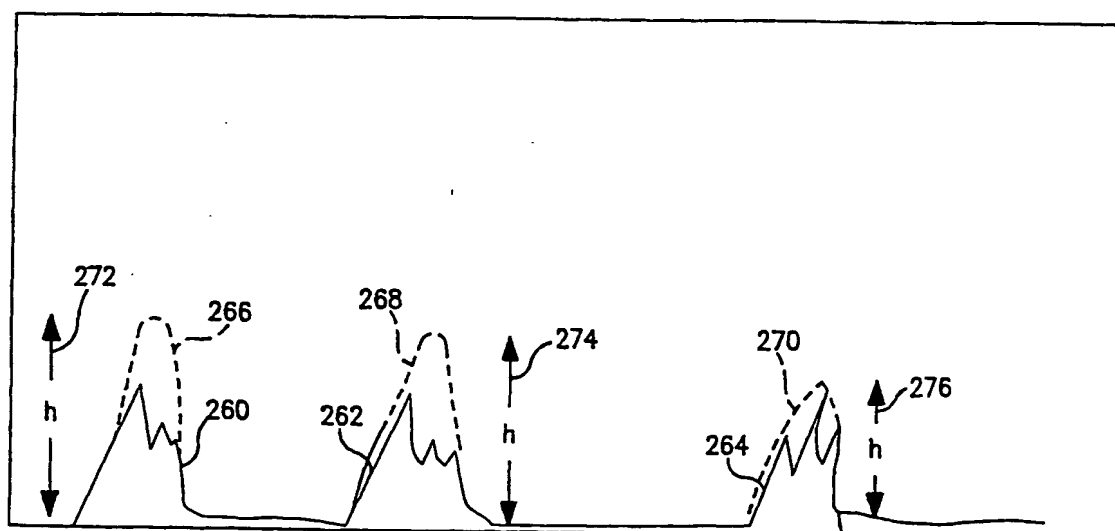


FIG. 49

SUBSTITUTE SHEET (RULE 26)

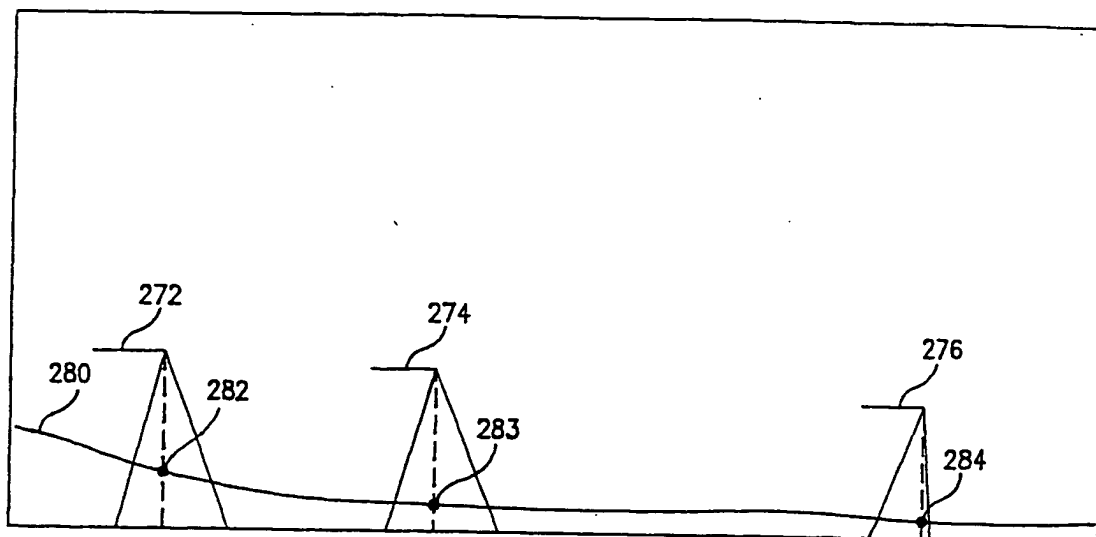


FIG. 50

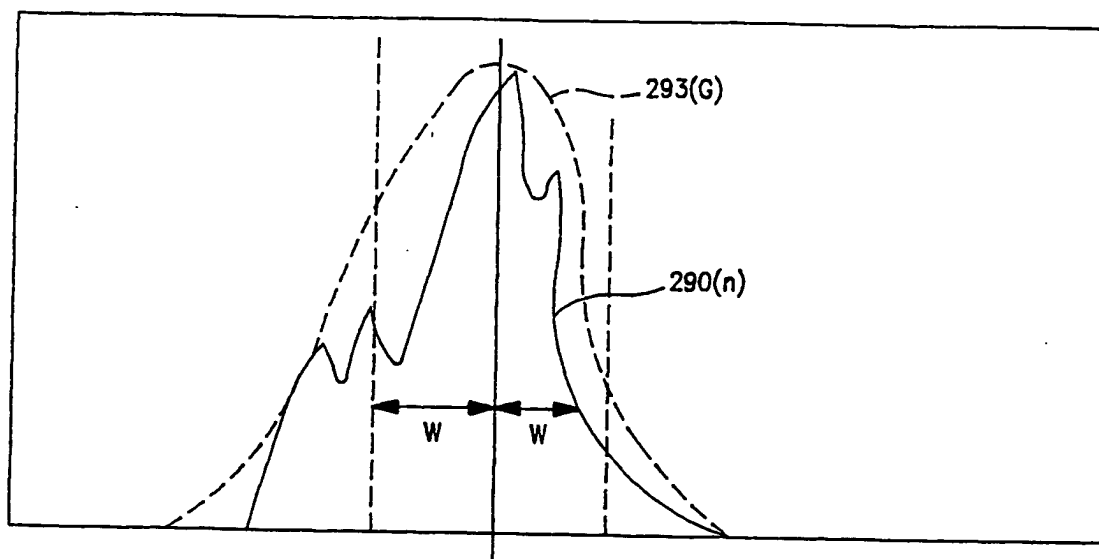


FIG. 51

SUBSTITUTE SHEET (RULE 26)

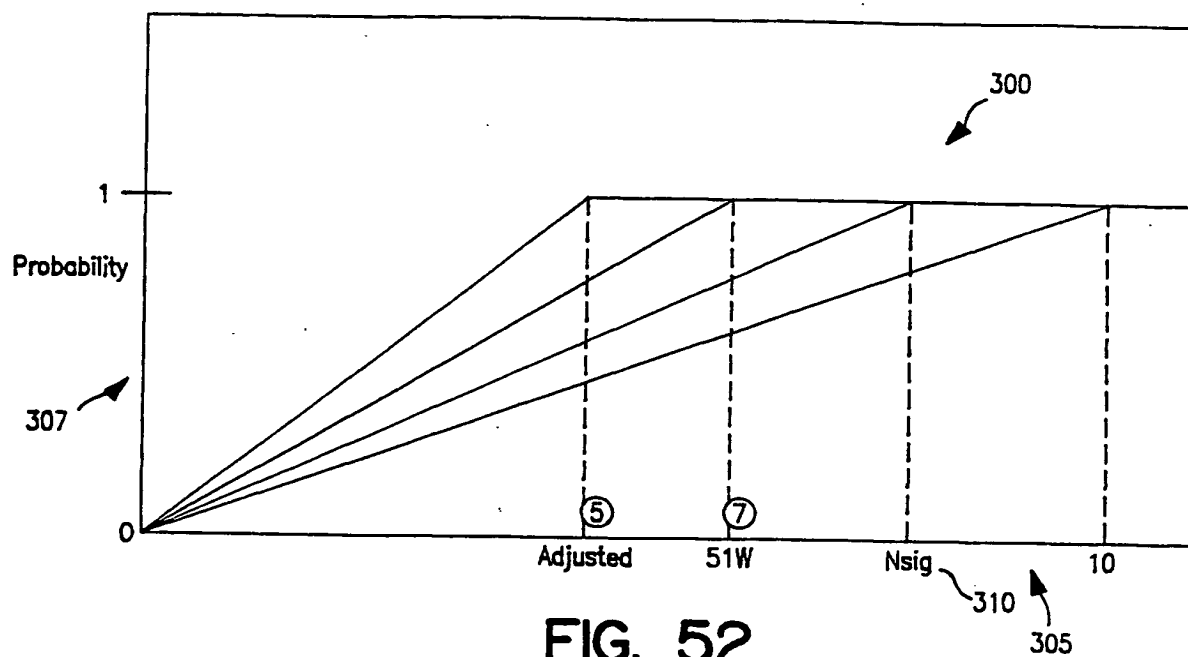


FIG. 52

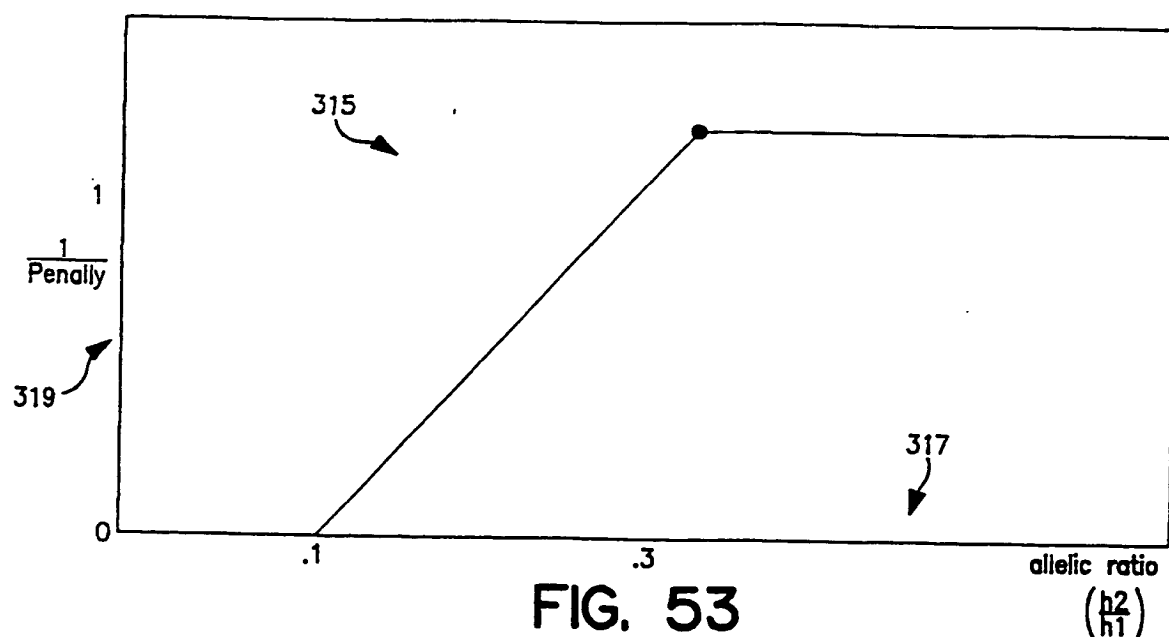


FIG. 53

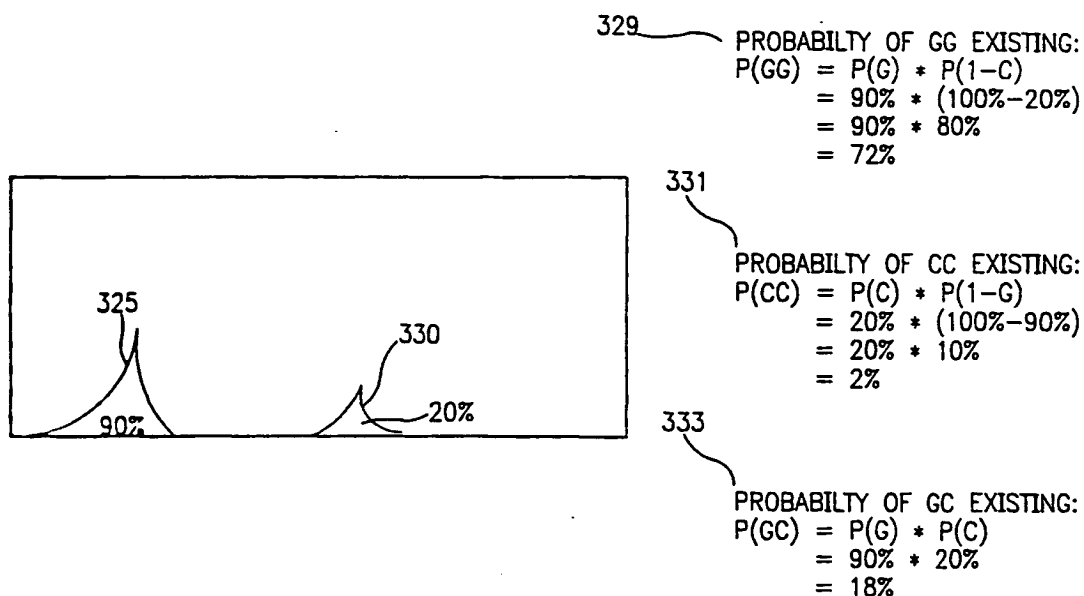


FIG. 54

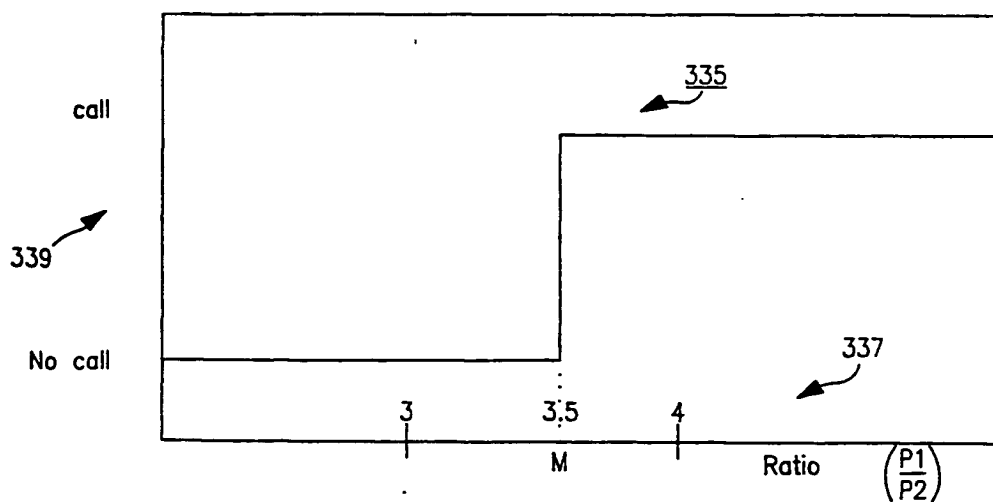


FIG. 55

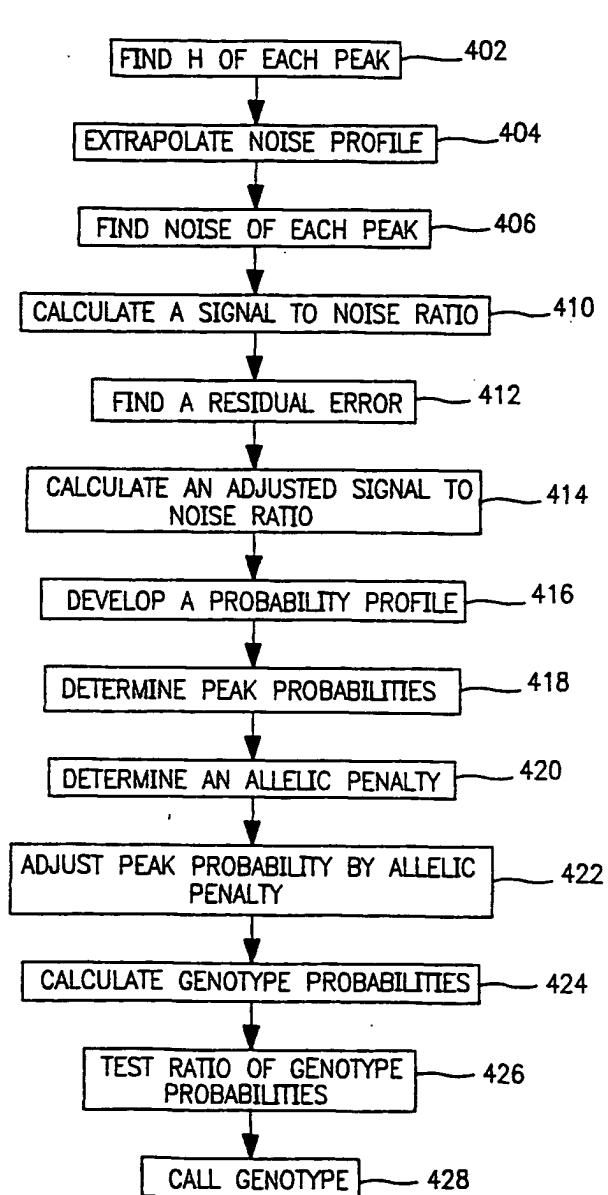


FIG. 56

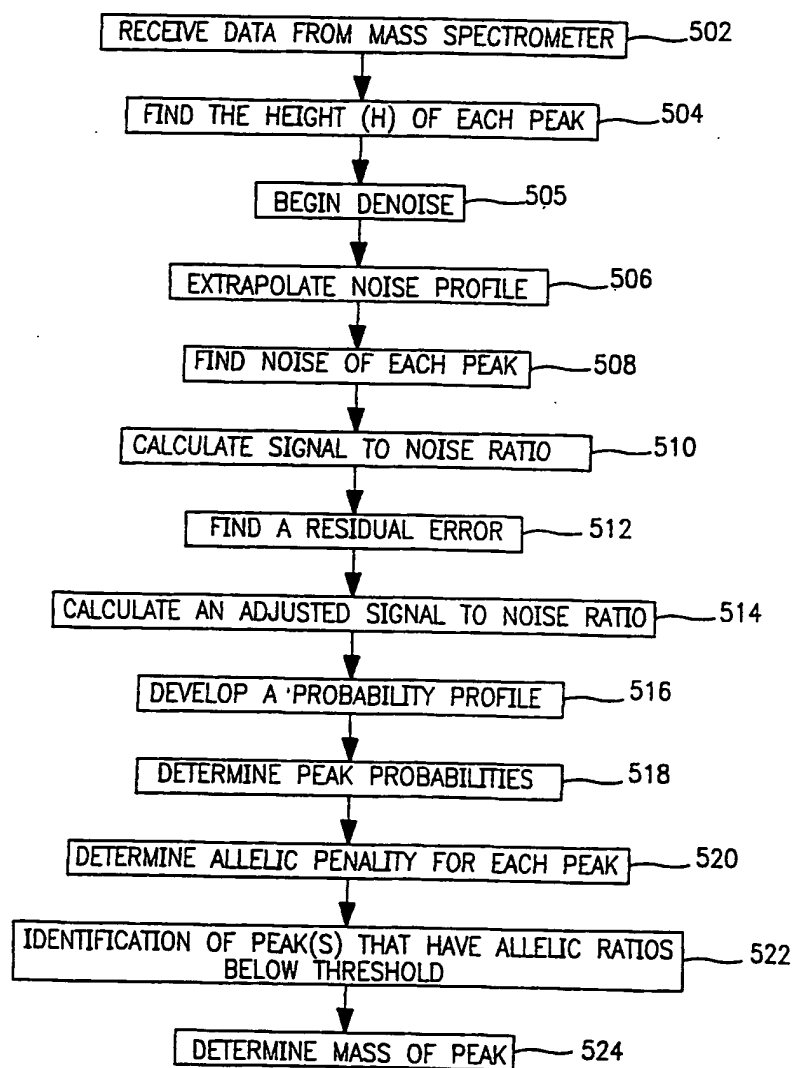


FIG. 57

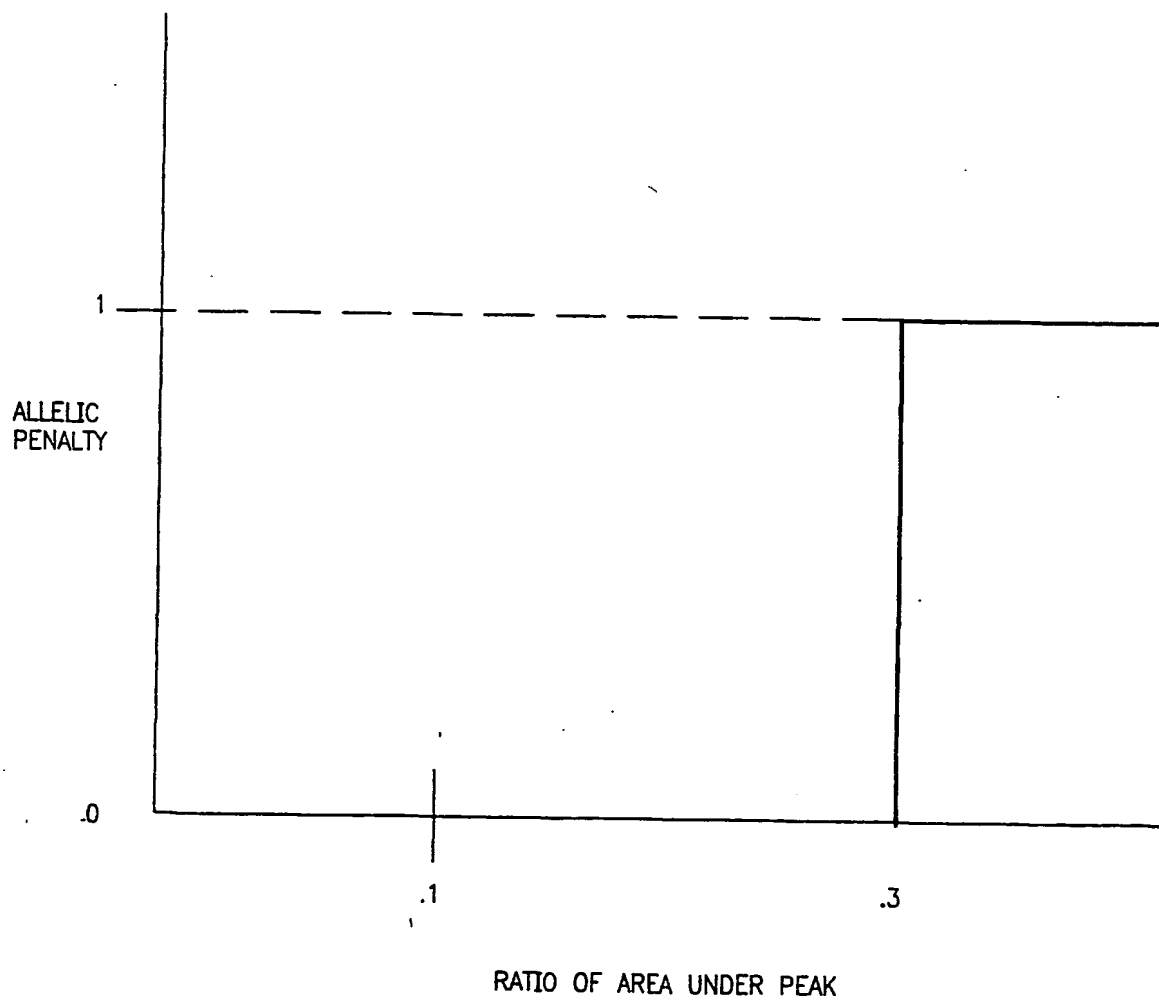


FIG. 58

## SEQUENCE LISTING

<110> SEQUENOM  
Braun et al.

<120> METHODS FOR GENERATING DATABASES AND DATABASES FOR IDENTIFYING  
POLYMORPHIC GENETIC MARKERS

<130> 24736-2033PC

<140> Not Yet Assigned

<141> 2000-10-13

<150> 60/217,658

<151> 2000-07-10

<150> 60/159,176

<151> 1999-10-13

<150> 60/217,251

<151> 2000-07-10

<150> 09/663,968

<151> 2000-09-19

<160> 118

<170> FastSeq for Windows Version 4.0

<210> 1

<211> 361

<212> DNA

<213> Homo Sapien

<400> 1

ctgaggacct	ggctcctctga	ctgctctttt	cacccatcta	cagtccccct	tgccgtccca	60
agcaatggat	gatttgatgc	tgtccccgga	cgatattgaa	caatgggtca	ctgaagaccc	120
aggtccagat	gaagctccca	gaatgccaga	ggctgctccc	cgcgtggccc	ctgcaccagc	180
agctcctaca	ccggcggccc	ctgcaccagc	ccccctctgg	cccctgtcat	cttctgtccc	240
ttcccagaaa	acctaccagg	gcagctacgg	tttccgtctg	ggcttcttgc	attctgggac	300
agccaagtct	gtgacttgca	cggtcagttg	ccctgagggg	ctggcttcca	tgagacttca	360
a						361

<210> 2

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide Primer

<400> 2

cccagtcacg	acgttgtaaa	acgctgagga	cctggctctc	tgac	44
------------	------------	------------	------------	------	----

<210> 3

<211> 42

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide Primer

<400> 3

agcggataac	aatttcacac	aggttgaagt	ctcatggaag	cc	42
------------	------------	------------	------------	----	----

<210> 4

<211> 17

<212> DNA

2/122

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 4

gccagaggct gctcccc

17

&lt;210&gt; 5

&lt;211&gt; 17

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 5

gccagaggct gctcccc

17

&lt;210&gt; 6

&lt;211&gt; 19

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 6

gccagaggct gctccccgc

19

&lt;210&gt; 7

&lt;211&gt; 18

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 7

gccagaggct gctcccc

18

&lt;210&gt; 8

&lt;211&gt; 161

&lt;212&gt; DNA

&lt;213&gt; Homo Sapien

&lt;400&gt; 8

gtccgtcaga acccatgcgg cagcaaggcc tgccgcgcgc tcttcggccc agtggacagc  
 gagcagctga gccgcgactg tgatgcgcta atggcgggct gcatccagga ggcccgtgag  
 cgatggaact tcgactttgt caccgagaca ccactggagg g

60

120

161

&lt;210&gt; 9

&lt;211&gt; 43

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligonucleotide Primer

&lt;400&gt; 9

3/122

cccagtcacg acgttgtaaa acggtccgtc agaaccatg cgg 43

<210> 10  
<211> 44  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 10  
agcggataac aatttcacac aggctccagt ggtgtctcgg tgac 44

<210> 11  
<211> 15  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Oligonucleotide Primer

<400> 11  
cagcgagcag ctgag 15

<210> 12  
<211> 15  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 12  
cagcgagcag ctgag 15

<210> 13  
<211> 16  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 13  
cagcgagcag ctgagc 16

<210> 14  
<211> 17  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 14  
cagcgagcag ctgagac 17

<210> 15  
<211> 205  
<212> DNA

4/122

&lt;213&gt; Homo Sapien

&lt;400&gt; 15

gcgctccatt	catctcttca	tcgactctct	gttgaatgaa	gaaaatccaa	gtaaggccta	60
caggtgcagt	tccaaggaag	cctttgagaa	agggctctgc	ttgagttgta	gaaagaaccg	120
ctgcaacaat	ctgggctatg	agatcaataa	agtcagagcc	aaaagaagca	gcaaaatgta	180
cctgaagact	cgttctcaga	tgccc				205

&lt;210&gt; 16

&lt;211&gt; 42

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligonucleotide Primers

&lt;400&gt; 16

cccagtcacg	acgttgtaaa	acggcgctcc	attcatctct	tc	42
------------	------------	------------	------------	----	----

&lt;210&gt; 17

&lt;211&gt; 42

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligonucleotide Primer

&lt;400&gt; 17

agcggataac	aatttcacac	aggggggcac	tgagaacgag	tc	42
------------	------------	------------	------------	----	----

&lt;210&gt; 18

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligonucleotide Primer

&lt;400&gt; 18

caatctgggc	tatgagatca	20
------------	------------	----

&lt;210&gt; 19

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

&lt;400&gt; 19

caatctgggc	tatgagatca	20
------------	------------	----

&lt;210&gt; 20

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Probe

5/122

<400> 20  
 caatctgggc tatgagatca a 21  
 <210> 21  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Probe  
 <400> 21  
 caatctgggc tatgagatca gt 20  
 <210> 22  
 <211> 60  
 <212> DNA  
 <213> Homo Sapien  
 <220>  
 <223> Probe  
 <400> 22  
 gtgccggcta ctccgatggc agcaaggact cctgcaaggg ggacagtgga ggcccacatg 60  
 <210> 23  
 <211> 60  
 <212> DNA  
 <213> Homo sapien  
 <400> 23  
 ccaccacta ccggggcacg tggtagctga cgggcatcgt cagctggggc cagggctgcg 60  
 <210> 24  
 <211> 42  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Oligonucleotide primer  
 <400> 24  
 cccagtcacg acgttgtaaa acgatggcag caaggactcc tg 42  
 <210> 25  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Oligonucleotide primer  
 <400> 25  
 cacatgccac ccactacc 18  
 <210> 26  
 <211> 43  
 <212> DNA  
 <213> Artificial Sequence

6/122

<220>  
<223> Oligonucleotide primer

<400> 26  
agcggataac aatttcacac aggtgacgat gcccgtcagg tac 43

<210> 27  
<211> 15  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 27  
atgccaccca ctacc 15

<210> 28  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 28  
cacatgccac ccactaccg 19

<210> 29  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 29  
cacatgccac ccactaccag 20

<210> 30  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Probe

<400> 30  
agcggataac aatttcacac agg 23

<210> 31  
<211> 2363  
<212> DNA  
<213> Homo Sapien

<220>  
<221> CDS  
<222> (138)... (2126)  
<223> AKAP-10

7/122

&lt;300&gt;

&lt;308&gt; GenBank AF037439

&lt;309&gt; 1997-12-21

&lt;400&gt; 31

```

gcggcttggt gataatatgg cggctggagc tgcctgggca tcccaggagg gcggtggggc      60
ccactcccgg aagaagggtc ccttttcgcg ctagtgcagc ggcccctctg gacccggaag      120
tccgggccgg ttgctga atg agg gga gcc ggg ccc tcc ccg cgc cag tcc      170
               Met Arg Gly Ala Gly Pro Ser Pro Arg Gln Ser
                   1                   5                   10

ccc cgc acc ctc cgt ccc gac ccg ggc ccc gcc atg tcc ttc ttc cgg      218
Pro Arg Thr Leu Arg Pro Asp Pro Gly Pro Ala Met Ser Phe Phe Arg
                   15                   20                   25

cgg aaa gtg aaa ggc aaa gaa caa gag aag acc tca gat gtg aag tcc      266
Arg Lys Val Lys Gly Lys Glu Gln Glu Lys Thr Ser Asp Val Lys Ser
                   30                   35                   40

att aaa gct tca ata tcc gta cat tcc cca caa aaa agc act aaa aat      314
Ile Lys Ala Ser Ile Ser Val His Ser Pro Gln Lys Ser Thr Lys Asn
                   45                   50                   55

cat gcc ttg ctg gag gct gca gga cca agt cat gtt gca atc aat gcc      362
His Ala Leu Leu Glu Ala Ala Gly Pro Ser His Val Ala Ile Asn Ala
                   60                   65                   70                   75

att tct gcc aac atg gac tcc ttt tca agt agc agg aca gcc aca ctt      410
Ile Ser Ala Asn Met Asp Ser Phe Ser Ser Ser Arg Thr Ala Thr Leu
                   80                   85                   90

aag aag cag cca agc cac atg gag gct gct cat ttt ggt gac ctg ggc      458
Lys Lys Gln Pro Ser His Met Glu Ala Ala His Phe Gly Asp Leu Gly
                   95                   100                   105

aga tct tgt ctg gac tac cag act caa gag acc aaa tca agc ctt tct      506
Arg Ser Cys Leu Asp Tyr Gln Thr Gln Glu Thr Lys Ser Ser Leu Ser
                   110                   115                   120

aag acc ctt gaa caa gtc ttg cac gac act att gtc ctc cct tac ttc      554
Lys Thr Leu Glu Gln Val Leu His Asp Thr Ile Val Leu Pro Tyr Phe
                   125                   130                   135

att caa ttc atg gaa ctt cgg cga atg gag cat ttg gtg aaa ttt tgg      602
Ile Gln Phe Met Glu Leu Arg Arg Met Glu His Leu Val Lys Phe Trp
                   140                   145                   150                   155

tta gag gct gaa agt ttt cat tca aca act tgg tcg cga ata aga gca      650
Leu Glu Ala Glu Ser Phe His Ser Thr Thr Trp Ser Arg Ile Arg Ala
                   160                   165                   170

cac agt cta aac aca atg aag cag agc tca ctg gct gag cct gtc tct      698
His Ser Leu Asn Thr Met Lys Gln Ser Ser Leu Ala Glu Pro Val Ser
                   175                   180                   185

cca tct aaa aag cat gaa act aca gcg tct ttt tta act gat tct ctt      746
Pro Ser Lys Lys His Glu Thr Thr Ala Ser Phe Leu Thr Asp Ser Leu
                   190                   195                   200

gat aag aga ttg gag gat tct ggc tca gca cag ttg ttt atg act cat      794

```

8/122

Asp	Lys	Arg	Leu	Glu	Asp	Ser	Gly	Ser	Ala	Gln	Leu	Phe	Met	Thr	His		
205						210					215						
tca	gaa	gga	att	gac	ctg	aat	aat	aga	act	aac	agc	act	cag	aat	cac		842
Ser	Glu	Gly	Ile	Asp	Leu	Asn	Asn	Arg	Thr	Asn	Ser	Thr	Gln	Asn	His		
220					225					230					235		
ttg	ctg	ctt	tcc	cag	gaa	tgt	gac	agt	gcc	cat	tct	ctc	cgt	ctt	gaa		890
Leu	Leu	Leu	Ser	Gln	Glu	Cys	Asp	Ser	Ala	His	Ser	Leu	Arg	Leu	Glu		
				240					245					250			
atg	gcc	aga	gca	gga	act	cac	caa	gtt	tcc	atg	gaa	acc	caa	gaa	tct		938
Met	Ala	Arg	Ala	Gly	Thr	His	Gln	Val	Ser	Met	Glu	Thr	Gln	Glu	Ser		
			255					260					265				
tcc	tct	aca	ctt	aca	gta	gcc	agt	aga	aat	agt	ccc	gct	tct	cca	cta		986
Ser	Ser	Thr	Leu	Thr	Val	Ala	Ser	Arg	Asn	Ser	Pro	Ala	Ser	Pro	Leu		
			270				275					280					
aaa	gaa	ttg	tca	gga	aaa	cta	atg	aaa	agt	ata	gaa	caa	gat	gca	gtg		1034
Lys	Glu	Leu	Ser	Gly	Lys	Leu	Met	Lys	Ser	Ile	Glu	Gln	Asp	Ala	Val		
	285					290					295						
aat	act	ttt	acc	aaa	tat	ata	tct	cca	gat	gct	gct	aaa	cca	ata	cca		1082
Asn	Thr	Phe	Thr	Lys	Tyr	Ile	Ser	Pro	Asp	Ala	Ala	Lys	Pro	Ile	Pro		
300					305					310					315		
att	aca	gaa	gca	atg	aga	aat	gac	atc	ata	gca	agg	att	tgt	gga	gaa		1130
Ile	Thr	Glu	Ala	Met	Arg	Asn	Asp	Ile	Ile	Ala	Arg	Ile	Cys	Gly	Glu		
				320					325					330			
gat	gga	cag	gtg	gat	ccc	aac	tgt	ttc	gtt	ttg	gca	cag	tcc	ata	gtc		1178
Asp	Gly	Gln	Val	Asp	Pro	Asn	Cys	Phe	Val	Leu	Ala	Gln	Ser	Ile	Val		
			335				340						345				
ttt	agt	gca	atg	gag	caa	gag	cac	ttt	agt	gag	ttt	ctg	cga	agt	cac		1226
Phe	Ser	Ala	Met	Glu	Gln	Glu	His	Phe	Ser	Glu	Phe	Leu	Arg	Ser	His		
		350					355					360					
cat	ttc	tgt	aaa	tac	cag	att	gaa	gtg	ctg	acc	agt	gga	act	gtt	tac		1274
His	Phe	Cys	Lys	Tyr	Gln	Ile	Glu	Val	Leu	Thr	Ser	Gly	Thr	Val	Tyr		
	365				370						375						
ctg	gct	gac	att	ctc	ttc	tgt	gag	tca	gcc	ctc	ttt	tat	ttc	tct	gag		1322
Leu	Ala	Asp	Ile	Leu	Phe	Cys	Glu	Ser	Ala	Leu	Phe	Tyr	Phe	Ser	Glu		
380					385					390					395		
tac	atg	gaa	aaa	gag	gat	gca	gtg	aat	atc	tta	caa	ttc	tgg	ttg	gca		1370
Tyr	Met	Glu	Lys	Glu	Asp	Ala	Val	Asn	Ile	Leu	Gln	Phe	Trp	Leu	Ala		
				400					405					410			
gca	gat	aac	ttc	cag	tct	cag	ctt	gct	gcc	aaa	aag	ggg	caa	tat	gat		1418
Ala	Asp	Asn	Phe	Gln	Ser	Gln	Leu	Ala	Ala	Lys	Lys	Gly	Gln	Tyr	Asp		
			415					420					425				
gga	cag	gag	gca	cag	aat	gat	gcc	atg	att	tta	tat	gac	aag	tac	ttc		1466
Gly	Gln	Glu	Ala	Gln	Asn	Asp	Ala	Met	Ile	Leu	Tyr	Asp	Lys	Tyr	Phe		
		430					435					440					
tcc	ctc	caa	gcc	aca	cat	cct	ctt	gga	ttt	gat	gat	gtt	gta	cga	tta		1514

9/122

Ser 445	Leu	Gln	Ala	Thr	His	Pro 450	Leu	Gly	Phe	Asp 455	Asp	Val	Val	Arg	Leu	
gaa 460	att	gaa	tcc	aat	atc	tgc	agg	gaa	ggt	ggg	cca	ctc	ccc	aac	tgt	1562
Glu	Ile	Glu	Ser	Asn	Ile	Cys	Arg	Glu	Gly	Gly	Pro	Leu	Pro	Asn	Cys	475
ttc	aca	act	cca	tta	cgt	cag	gcc	tgg	aca	acc	atg	gag	aag	gtc	ttt	1610
Phe	Thr	Thr	Pro	Leu	Arg	Gln	Ala	Trp	Thr	Thr	Met	Glu	Lys	Val	Phe	490
ttg	cct	ggc	ttt	ctg	tcc	agc	aat	ctt	tat	tat	aaa	tat	ttg	aat	gat	1658
Leu	Pro	Gly	Phe	Leu	Ser	Ser	Asn	Leu	Tyr	Tyr	Lys	Tyr	Leu	Asn	Asp	505
ctc	atc	cat	tgc	gtt	cga	gga	gat	gaa	ttt	ctg	ggc	ggg	aac	gtg	tgc	1706
Leu	Ile	His	Ser	Val	Arg	Gly	Asp	Glu	Phe	Leu	Gly	Gly	Asn	Val	Ser	520
ccg	act	ggt	cct	ggc	tct	gtt	ggc	cct	cct	gat	gag	tct	cac	cca	ggg	1754
Pro	Thr	Ala	Pro	Gly	Ser	Val	Gly	Pro	Pro	Asp	Glu	Ser	His	Pro	Gly	535
agt	tct	gac	agc	tct	gcg	tct	cag	tcc	agt	gtg	aaa	aaa	gcc	agt	att	1802
Ser	Ser	Asp	Ser	Ser	Ala	Ser	Gln	Ser	Ser	Val	Lys	Lys	Ala	Ser	Ile	555
aaa	ata	ctg	aaa	aat	ttt	gat	gaa	gcg	ata	att	gtg	gat	gcg	gca	agt	1850
Lys	Ile	Leu	Lys	Asn	Phe	Asp	Glu	Ala	Ile	Ile	Val	Asp	Ala	Ala	Ser	570
ctg	gat	cca	gaa	tct	tta	tat	caa	cgg	aca	tat	gcc	ggg	aag	atg	aca	1898
Leu	Asp	Pro	Glu	Ser	Leu	Tyr	Gln	Arg	Thr	Tyr	Ala	Gly	Lys	Met	Thr	585
ttt	gga	aga	gtg	agt	gac	ttg	ggg	caa	ttc	atc	cgg	gaa	tct	gag	cct	1946
Phe	Gly	Arg	Val	Ser	Asp	Leu	Gly	Gln	Phe	Ile	Arg	Glu	Ser	Glu	Pro	600
gaa	cct	gat	gta	agg	aaa	tca	aaa	gga	tcc	atg	ttc	tca	caa	gct	atg	1994
Glu	Pro	Asp	Val	Arg	Lys	Ser	Lys	Gly	Ser	Met	Phe	Ser	Gln	Ala	Met	615
aag	aaa	tgg	gtg	caa	gga	aat	act	gat	gag	gcc	cag	gaa	gag	cta	gct	2042
Lys	Lys	Trp	Val	Gln	Gly	Asn	Thr	Asp	Glu	Ala	Gln	Glu	Glu	Leu	Ala	635
tgg	aag	att	gct	aaa	atg	ata	gtc	agt	gac	att	atg	cag	cag	gct	cag	2090
Trp	Lys	Ile	Ala	Lys	Met	Ile	Val	Ser	Asp	Ile	Met	Gln	Gln	Ala	Gln	650
tat	gat	caa	ccg	tta	gag	aaa	tct	aca	aag	tta	tga	ctcaaaactt				2136
Tyr	Asp	Gln	Pro	Leu	Glu	Lys	Ser	Thr	Lys	Leu	*					655
gagataaagg	aaatctgctt	gtgaaaaata	agagaacttt	tttcccttgg	ttggattctt											2196
caacacagcc	aatgaaaaca	gcactatatt	tctgatctgt	cactgttggt	tccagggaga											2256
gaatggggag	acaatcctag	gacttccacc	ctaatagcagt	tacctgtagg	gcataattgg											2316
atggcacatg	atgtttcaca	cagtgaggag	tctttaaagg	ttaccaa												2363

10/122

<210> 32  
 <211> 662  
 <212> PRT  
 <213> Homo Sapien

<400> 32  
 Met Arg Gly Ala Gly Pro Ser Pro Arg Gln Ser Pro Arg Thr Leu Arg  
 1 5 10 15  
 Pro Asp Pro Gly Pro Ala Met Ser Phe Phe Arg Arg Lys Val Lys Gly  
 20 25 30  
 Lys Glu Gln Glu Lys Thr Ser Asp Val Lys Ser Ile Lys Ala Ser Ile  
 35 40 45  
 Ser Val His Ser Pro Gln Lys Ser Thr Lys Asn His Ala Leu Leu Glu  
 50 55 60  
 Ala Ala Gly Pro Ser His Val Ala Ile Asn Ala Ile Ser Ala Asn Met  
 65 70 75 80  
 Asp Ser Phe Ser Ser Ser Arg Thr Ala Thr Leu Lys Lys Gln Pro Ser  
 85 90 95  
 His Met Glu Ala Ala His Phe Gly Asp Leu Gly Arg Ser Cys Leu Asp  
 100 105 110  
 Tyr Gln Thr Gln Glu Thr Lys Ser Ser Leu Ser Lys Thr Leu Glu Gln  
 115 120 125  
 Val Leu His Asp Thr Ile Val Leu Pro Tyr Phe Ile Gln Phe Met Glu  
 130 135 140  
 Leu Arg Arg Met Glu His Leu Val Lys Phe Trp Leu Glu Ala Glu Ser  
 145 150 155 160  
 Phe His Ser Thr Thr Trp Ser Arg Ile Arg Ala His Ser Leu Asn Thr  
 165 170 175  
 Met Lys Gln Ser Ser Leu Ala Glu Pro Val Ser Pro Ser Lys Lys His  
 180 185 190  
 Glu Thr Thr Ala Ser Phe Leu Thr Asp Ser Leu Asp Lys Arg Leu Glu  
 195 200 205  
 Asp Ser Gly Ser Ala Gln Leu Phe Met Thr His Ser Glu Gly Ile Asp  
 210 215 220  
 Leu Asn Asn Arg Thr Asn Ser Thr Gln Asn His Leu Leu Leu Ser Gln  
 225 230 235 240  
 Glu Cys Asp Ser Ala His Ser Leu Arg Leu Glu Met Ala Arg Ala Gly  
 245 250 255  
 Thr His Gln Val Ser Met Glu Thr Gln Glu Ser Ser Ser Thr Leu Thr  
 260 265 270  
 Val Ala Ser Arg Asn Ser Pro Ala Ser Pro Leu Lys Glu Leu Ser Gly  
 275 280 285  
 Lys Leu Met Lys Ser Ile Glu Gln Asp Ala Val Asn Thr Phe Thr Lys  
 290 295 300  
 Tyr Ile Ser Pro Asp Ala Ala Lys Pro Ile Pro Ile Thr Glu Ala Met  
 305 310 315 320  
 Arg Asn Asp Ile Ile Ala Arg Ile Cys Gly Glu Asp Gly Gln Val Asp  
 325 330 335  
 Pro Asn Cys Phe Val Leu Ala Gln Ser Ile Val Phe Ser Ala Met Glu  
 340 345 350  
 Gln Glu His Phe Ser Glu Phe Leu Arg Ser His His Phe Cys Lys Tyr  
 355 360 365  
 Gln Ile Glu Val Leu Thr Ser Gly Thr Val Tyr Leu Ala Asp Ile Leu  
 370 375 380  
 Phe Cys Glu Ser Ala Leu Phe Tyr Phe Ser Glu Tyr Met Glu Lys Glu  
 385 390 395 400  
 Asp Ala Val Asn Ile Leu Gln Phe Trp Leu Ala Ala Asp Asn Phe Gln  
 405 410 415  
 Ser Gln Leu Ala Lys Lys Gly Gln Tyr Asp Gly Gln Glu Ala Gln  
 420 425 430

11/122

Asn Asp Ala Met Ile Leu Tyr Asp Lys Tyr Phe Ser Leu Gln Ala Thr  
 435 440 445  
 His Pro Leu Gly Phe Asp Asp Val Val Arg Leu Glu Ile Glu Ser Asn  
 450 455 460  
 Ile Cys Arg Glu Gly Gly Pro Leu Pro Asn Cys Phe Thr Thr Pro Leu  
 465 470 475 480  
 Arg Gln Ala Trp Thr Thr Met Glu Lys Val Phe Leu Pro Gly Phe Leu  
 485 490 495  
 Ser Ser Asn Leu Tyr Tyr Lys Tyr Leu Asn Asp Leu Ile His Ser Val  
 500 505 510  
 Arg Gly Asp Glu Phe Leu Gly Gly Asn Val Ser Pro Thr Ala Pro Gly  
 515 520 525  
 Ser Val Gly Pro Pro Asp Glu Ser His Pro Gly Ser Ser Asp Ser Ser  
 530 535 540  
 Ala Ser Gln Ser Ser Val Lys Lys Ala Ser Ile Lys Ile Leu Lys Asn  
 545 550 555 560  
 Phe Asp Glu Ala Ile Ile Val Asp Ala Ala Ser Leu Asp Pro Glu Ser  
 565 570 575  
 Leu Tyr Gln Arg Thr Tyr Ala Gly Lys Met Thr Phe Gly Arg Val Ser  
 580 585 590  
 Asp Leu Gly Gln Phe Ile Arg Glu Ser Glu Pro Glu Pro Asp Val Arg  
 595 600 605  
 Lys Ser Lys Gly Ser Met Phe Ser Gln Ala Met Lys Lys Trp Val Gln  
 610 615 620  
 Gly Asn Thr Asp Glu Ala Gln Glu Glu Leu Ala Trp Lys Ile Ala Lys  
 625 630 635 640  
 Met Ile Val Ser Asp Ile Met Gln Gln Ala Gln Tyr Asp Gln Pro Leu  
 645 650 655  
 Glu Lys Ser Thr Lys Leu  
 660

&lt;210&gt; 33

&lt;211&gt; 2363

&lt;212&gt; DNA

&lt;213&gt; Homo Sapien

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (138)...(2126)

&lt;223&gt; AKAP-10-5

&lt;221&gt; allele

&lt;222&gt; 2073

&lt;223&gt; Single Nucleotide Polymorphism: A to G

&lt;400&gt; 33

gcggcttggt gataatatgg cggctggagc tgcctgggca tcccaggagg gcggtggggc 60  
 ccactcccgg aagaagggtc ccttttcgcg ctagtgcagc ggcccctctg gacccggaag 120  
 tccgggcccgg ttgctga atg agg gga gcc ggg ccc tcc ccg cgc cag tcc 170  
 Met Arg Gly Ala Gly Pro Ser Pro Arg Gln Ser  
 1 5 10  
 ccc cgc acc ctc cgt ccc gac ccg ggc ccc gcc atg tcc ttc ttc cgg 218  
 Pro Arg Thr Leu Arg Pro Asp Pro Gly Pro Ala Met Ser Phe Phe Arg  
 15 20 25  
 cgg aaa gtg aaa ggc aaa gaa caa gag aag acc tca gat gtg aag tcc 266  
 Arg Lys Val Lys Gly Lys Glu Gln Glu Lys Thr Ser Asp Val Lys Ser  
 30 35 40

12/122

att aaa gct tca ata tcc gta cat tcc cca caa aaa agc act aaa aat Ile Lys Ala Ser Ile Ser Val His Ser Pro Gln Lys Ser Thr Lys Asn 45 50 55	314
cat gcc ttg ctg gag gct gca gga cca agt cat gtt gca atc aat gcc His Ala Leu Leu Glu Ala Ala Gly Pro Ser His Val Ala Ile Asn Ala 60 65 70 75	362
att tct gcc aac atg gac tcc ttt tca agt agc agg aca gcc aca ctt Ile Ser Ala Asn Met Asp Ser Phe Ser Ser Ser Arg Thr Ala Thr Leu 80 85 90	410
aag aag cag cca agc cac atg gag gct gct cat ttt ggt gac ctg ggc Lys Lys Gln Pro Ser His Met Glu Ala Ala His Phe Gly Asp Leu Gly 95 100 105	458
agt tct tgt ctg gac tac cag act caa gag acc aaa tca agc ctt tct Arg Ser Cys Leu Asp Tyr Gln Thr Gln Glu Thr Lys Ser Ser Leu Ser 110 115 120	506
aag acc ctt gaa caa gtc ttg cac gac act att gtc ctc cct tac ttc Lys Thr Leu Glu Gln Val Leu His Asp Thr Ile Val Leu Pro Tyr Phe 125 130 135	554
att caa ttc atg gaa ctt cgg cga atg gag cat ttg gtg aaa ttt tgg Ile Gln Phe Met Glu Leu Arg Arg Met Glu His Leu Val Lys Phe Trp 140 145 150 155	602
tta gag gct gaa agt ttt cat tca aca act tgg tct cga ata aga gca Leu Glu Ala Glu Ser Phe His Ser Thr Thr Trp Ser Arg Ile Arg Ala 160 165 170	650
cac agt cta aac aca atg aag cag agc tca ctg gct gag cct gtc tct His Ser Leu Asn Thr Met Lys Gln Ser Ser Leu Ala Glu Pro Val Ser 175 180 185	698
cca tct aaa aag cat gaa act aca gcg tct ttt tta act gat tct ctt Pro Ser Lys Lys His Glu Thr Thr Ala Ser Phe Leu Thr Asp Ser Leu 190 195 200	746
gat aag aga ttg gag gat tct ggc tca gca cag ttg ttt atg act cat Asp Lys Arg Leu Glu Asp Ser Gly Ser Ala Gln Leu Phe Met Thr His 205 210 215	794
tca gaa gga att gac ctg aat aat aga act aac agc act cag aat cac Ser Glu Gly Ile Asp Leu Asn Asn Arg Thr Asn Ser Thr Gln Asn His 220 225 230 235	842
ttg ctg ctt tcc cag gaa tgt gac agt gcc cat tct ctc cgt ctt gaa Leu Leu Leu Ser Gln Glu Cys Asp Ser Ala His Ser Leu Arg Leu Glu 240 245 250	890
atg gcc aga gca gga act cac caa gtt tcc atg gaa acc caa gaa tct Met Ala Arg Ala Gly Thr His Gln Val Ser Met Glu Thr Gln Glu Ser 255 260 265	938
tcc tct aca ctt aca gta gcc agt aga aat agt ccc gct tct cca cta Ser Ser Thr Leu Thr Val Ala Ser Arg Asn Ser Pro Ala Ser Pro Leu 270 275 280	986

13/122

aaa gaa ttg tca gga aaa cta atg aaa agt ata gaa caa gat gca gtg Lys Glu Leu Ser Gly Lys Leu Met Lys Ser Ile Glu Gln Asp Ala Val 285 290 295	1034
aat act ttt acc aaa tat ata tct cca gat gct gct aaa cca ata cca Asn Thr Phe Thr Lys Tyr Ile Ser Pro Asp Ala Ala Lys Pro Ile Pro 300 305 310 315	1082
att aca gaa gca atg aga aat gac atc ata gca agg att tgt gga gaa Ile Thr Glu Ala Met Arg Asn Asp Ile Ile Ala Arg Ile Cys Gly Glu 320 325 330	1130
gat gga cag gtg gat ccc aac tgt ttc gtt ttg gca cag tcc ata gtc Asp Gly Gln Val Asp Pro Asn Cys Phe Val Leu Ala Gln Ser Ile Val 335 340 345	1178
ttt agt gca atg gag caa gag cac ttt agt gag ttt ctg cga agt cac Phe Ser Ala Met Glu Gln Glu His Phe Ser Glu Phe Leu Arg Ser His 350 355 360	1226
cat ttc tgt aaa tac cag att gaa gtg ctg acc agt gga act gtt tac His Phe Cys Lys Tyr Gln Ile Glu Val Leu Thr Ser Gly Thr Val Tyr 365 370 375	1274
ctg gct gac att ctc ttc tgt gag tca gcc ctc ttt tat ttc tct gag Leu Ala Asp Ile Leu Phe Cys Glu Ser Ala Leu Phe Tyr Phe Ser Glu 380 385 390 395	1322
tac atg gaa aaa gag gat gca gtg aat atc tta caa ttc tgg ttg gca Tyr Met Glu Lys Glu Asp Ala Val Asn Ile Leu Gln Phe Trp Leu Ala 400 405 410	1370
gca gat aac ttc cag tct cag ctt gct gcc aaa aag ggg caa tat gat Ala Asp Asn Phe Gln Ser Gln Leu Ala Ala Lys Lys Gly Gln Tyr Asp 415 420 425	1418
gga cag gag gca cag aat gat gcc atg att tta tat gac aag tac ttc Gly Gln Glu Ala Gln Asn Asp Ala Met Ile Leu Tyr Asp Lys Tyr Phe 430 435 440	1466
tcc ctc caa gcc aca cat cct ctt gga ttt gat gat gtt gta cga tta Ser Leu Gln Ala Thr His Pro Leu Gly Phe Asp Asp Val Val Arg Leu 445 450 455	1514
gaa att gaa tcc aat atc tgc agg gaa ggt ggg cca ctc ccc aac tgt Glu Ile Glu Ser Asn Ile Cys Arg Glu Gly Gly Pro Leu Pro Asn Cys 460 465 470 475	1562
ttc aca act cca tta cgt cag gcc tgg aca acc atg gag aag gtc ttt Phe Thr Thr Pro Leu Arg Gln Ala Trp Thr Thr Met Glu Lys Val Phe 480 485 490	1610
ttg cct ggc ttt ctg tcc agc aat ctt tat tat aaa tat ttg aat gat Leu Pro Gly Phe Leu Ser Ser Asn Leu Tyr Tyr Lys Tyr Asn Asp 495 500 505	1658
ctc atc cat tcg gtt cga gga gat gaa ttt ctg ggc ggg aac gtg tcg Leu Ile His Ser Val Arg Gly Asp Glu Phe Leu Gly Gly Asn Val Ser 510 515 520	1706

14/122

ccg act gct cct ggc tct gtt ggc cct cct gat gag tct cac cca ggg 1754  
 Pro Thr Ala Pro Gly Ser Val Gly Pro Pro Asp Glu Ser His Pro Gly  
 525 530 535

agt tct gac agc tct gcg tct cag tcc agt gtg aaa aaa gcc agt att 1802  
 Ser Ser Asp Ser Ser Ala Ser Gln Ser Ser Val Lys Lys Ala Ser Ile  
 540 545 550 555

aaa ata ctg aaa aat ttt gat gaa gcg ata att gtg gat gcg gca agt 1850  
 Lys Ile Leu Lys Asn Phe Asp Glu Ala Ile Ile Val Asp Ala Ala Ser  
 560 565 570

ctg gat cca gaa tct tta tat caa cgg aca tat gcc ggg aag atg aca 1898  
 Leu Asp Pro Glu Ser Leu Tyr Gln Arg Thr Tyr Ala Gly Lys Met Thr  
 575 580 585

ttt gga aga gtg agt gac ttg ggg caa ttc atc cgg gaa tct gag cct 1946  
 Phe Gly Arg Val Ser Asp Leu Gly Gln Phe Ile Arg Glu Ser Glu Pro  
 590 595 600

gaa cct gat gta agg aaa tca aaa gga tcc atg ttc tca caa gct atg 1994  
 Glu Pro Asp Val Arg Lys Ser Lys Gly Ser Met Phe Ser Gln Ala Met  
 605 610 615

aag aaa tgg gtg caa gga aat act gat gag gcc cag gaa gag cta gct 2042  
 Lys Lys Trp Val Gln Gly Asn Thr Asp Glu Ala Gln Glu Glu Leu Ala  
 620 625 630 635

tgg aag att gct aaa atg ata gtc agt gac gtt atg cag cag gct cag 2090  
 Trp Lys Ile Ala Lys Met Ile Val Ser Asp Val Met Gln Gln Ala Gln  
 640 645 650

tat gat caa ccg tta gag aaa tct aca aag tta tga ctcaaaactt 2136  
 Tyr Asp Gln Pro Leu Glu Lys Ser Thr Lys Leu \*  
 655 660

gagataaagg aaatctgctt gtgaaaaata agagaacttt tttcccttgg ttggattcctt 2196  
 caacacagcc aatgaaaaca gcactatatt tctgatctgt cactgttggt tccagggaga 2256  
 gaatggggag acaatcctag gacttccacc ctaatgcagt taccgttagg gcataattgg 2316  
 atggcacatg atgtttcaca cagtgaggag tctttaaagg ttaccaa 2363

<210> 34  
 <211> 662  
 <212> PRT  
 <213> Homo Sapien

<400> 34  
 Met Arg Gly Ala Gly Pro Ser Pro Arg Gln Ser Pro Arg Thr Leu Arg  
 1 5 10 15  
 Pro Asp Pro Gly Pro Ala Met Ser Phe Arg Arg Lys Val Lys Gly  
 20 25 30  
 Lys Glu Gln Glu Lys Thr Ser Asp Val Lys Ser Ile Lys Ala Ser Ile  
 35 40 45  
 Ser Val His Ser Pro Gln Lys Ser Thr Lys Asn His Ala Leu Leu Glu  
 50 55 60  
 Ala Ala Gly Pro Ser His Val Ala Ile Asn Ala Ile Ser Ala Asn Met  
 65 70 75 80  
 Asp Ser Phe Ser Ser Arg Thr Ala Thr Leu Lys Lys Gln Pro Ser  
 85 90 95  
 His Met Glu Ala Ala His Phe Gly Asp Leu Gly Arg Ser Cys Leu Asp

				100					105					110		
Tyr	Gln	Thr	Gln	Glu	Thr	Lys	Ser	Ser	Leu	Ser	Lys	Thr	Leu	Glu	Gln	
		115					120					125				
Val	Leu	His	Asp	Thr	Ile	Val	Leu	Pro	Tyr	Phe	Ile	Gln	Phe	Met	Glu	
	130					135					140					
Leu	Arg	Arg	Met	Glu	His	Leu	Val	Lys	Phe	Trp	Leu	Glu	Ala	Glu	Ser	
145				150					155						160	
Phe	His	Ser	Thr	Thr	Trp	Ser	Arg	Ile	Arg	Ala	His	Ser	Leu	Asn	Thr	
			165					170						175		
Met	Lys	Gln	Ser	Ser	Leu	Ala	Glu	Pro	Val	Ser	Pro	Ser	Lys	Lys	His	
		180					185					190				
Glu	Thr	Thr	Ala	Ser	Phe	Leu	Thr	Asp	Ser	Leu	Asp	Lys	Arg	Leu	Glu	
	195						200					205				
Asp	Ser	Gly	Ser	Ala	Gln	Leu	Phe	Met	Thr	His	Ser	Glu	Gly	Ile	Asp	
	210			215							220					
Leu	Asn	Asn	Arg	Thr	Asn	Ser	Thr	Gln	Asn	His	Leu	Leu	Leu	Ser	Gln	
225				230					235						240	
Glu	Cys	Asp	Ser	Ala	His	Ser	Leu	Arg	Leu	Glu	Met	Ala	Arg	Ala	Gly	
			245					250						255		
Thr	His	Gln	Val	Ser	Met	Glu	Thr	Gln	Glu	Ser	Ser	Ser	Thr	Leu	Thr	
		260						265					270			
Val	Ala	Ser	Arg	Asn	Ser	Pro	Ala	Ser	Pro	Leu	Lys	Glu	Leu	Ser	Gly	
	275						280					285				
Lys	Leu	Met	Lys	Ser	Ile	Glu	Gln	Asp	Ala	Val	Asn	Thr	Phe	Thr	Lys	
	290			295							300					
Tyr	Ile	Ser	Pro	Asp	Ala	Ala	Lys	Pro	Ile	Pro	Ile	Thr	Glu	Ala	Met	
305				310						315					320	
Arg	Asn	Asp	Ile	Ile	Ala	Arg	Ile	Cys	Gly	Glu	Asp	Gly	Gln	Val	Asp	
			325						330					335		
Pro	Asn	Cys	Phe	Val	Leu	Ala	Gln	Ser	Ile	Val	Phe	Ser	Ala	Met	Glu	
		340					345						350			
Gln	Glu	His	Phe	Ser	Glu	Phe	Leu	Arg	Ser	His	His	Phe	Cys	Lys	Tyr	
	355						360					365				
Gln	Ile	Glu	Val	Leu	Thr	Ser	Gly	Thr	Val	Tyr	Leu	Ala	Asp	Ile	Leu	
	370				375						380					
Phe	Cys	Glu	Ser	Ala	Leu	Phe	Tyr	Phe	Ser	Glu	Tyr	Met	Glu	Lys	Glu	
385				390						395					400	
Asp	Ala	Val	Asn	Ile	Leu	Gln	Phe	Trp	Leu	Ala	Ala	Asp	Asn	Phe	Gln	
			405						410					415		
Ser	Gln	Leu	Ala	Ala	Lys	Lys	Gly	Gln	Tyr	Asp	Gly	Gln	Glu	Ala	Gln	
		420						425					430			
Asn	Asp	Ala	Met	Ile	Leu	Tyr	Asp	Lys	Tyr	Phe	Ser	Leu	Gln	Ala	Thr	
	435						4									

16/122

Asp	Leu	Gly	Gln	Phe	Ile	Arg	Glu	585	Ser	Glu	Pro	Glu	Pro	590	Asp	Val	Arg
		595					600						605				
Lys	Ser	Lys	Gly	Ser	Met	Phe	Ser	Gln	Ala	Met	Lys	Lys	Trp	Val	Gln		
		610				615						620					
Gly	Asn	Thr	Asp	Glu	Ala	Gln	Glu	Glu	Leu	Ala	Trp	Lys	Ile	Ala	Lys		
625					630					635					640		
Met	Ile	Val	Ser	Asp	Val	Met	Gln	Gln	Ala	Gln	Tyr	Asp	Gln	Pro	Leu		
				645					650					655			
Glu	Lys	Ser	Thr	Lys	Leu												
			660														

&lt;210&gt; 35

&lt;211&gt; 162025

&lt;212&gt; DNA

&lt;213&gt; Homo Sapien

&lt;300&gt;

&lt;308&gt; GenBank AC005730

&lt;309&gt; 1998-10-22

&lt;400&gt; 35

gaattcctat	ttcaaaagaa	acaaatgggc	caagtatggg	ggctcatacc	tgtaatccca	60
gcactttggg	aggccgaggt	gagtggtgca	cttgaggtca	ggagttccag	gccagtctgg	120
ccaacatggt	gaaacactgt	ctctactaaa	aatacaaaaa	ttagccgggc	gtgggtggcg	180
gcacctgtaa	tcccagctac	tcaggagggt	gaggcaggag	aattgcttga	acctgggaga	240
tggagggtgc	agtgagccga	gatcgcgcca	ctgctctcca	gcctgggtgg	cagagtgaga	300
ctctgtctca	aaaagaaaaca	aagaaataaa	tgaaacaatt	ttgttcacat	atatttcaca	360
aatttgaaat	gttaaaggta	ttatgggtcac	tgatatcctg	tttcattctt	tatataatca	420
ttaagtttga	aatgtatact	tgcaactacta	acacagtagt	taatcttagt	cctacaagtt	480
actgctttta	cacaatatat	tttcgttaata	tgtatgcact	gggtgtttatg	tacgtgttta	540
tgtttatatc	tgttaaaatt	agcagtttcc	atctttttct	atctttgtacc	atcacatcag	600
ttcagaagga	ttgacagagc	aaaatgattt	gatgaagtat	aaaagtcaca	tggtgagtg	660
cataaataca	actctgaaca	attaggaggc	tcactattga	ctggaactaa	actgcaagcc	720
agaaagacac	atatectata	tgtaagagaa	tgtaccaccc	aggcaggtta	agaagggaag	780
tacacataga	aagcacaaatg	gtgaataaatt	aaaaaattgg	aatttatcag	acactggatt	840
catttgctcc	taaagtcaga	gtcctctatt	gtttttttgt	ttttgtgggt	ttctttttta	900
atctttttat	ttttttgtaga	gtcggaggtct	cactgtgtta	cccgggctgg	tctagaactc	960
ctggcctcaa	acaaacctcc	tgccctcagct	tcccaaagca	ttgggattac	agacatgagc	1020
cactgagccc	agcccagacg	ctttagcatt	tatgaagctt	ctgaaatagt	tgtagaaacc	1080
gcataagctt	tccatgtcac	tttcaaagtt	tgatggtctc	tttagtaaac	caaccaagtt	1140
attcctcaag	ggcaaaaataa	cattttctcag	tgcaaaaactg	atgcacttca	ttaccaaag	1200
gaaaagacca	caactataga	ggcgtcattg	aaagctgcac	tcttcagagg	ccaaaaaaa	1260
agggtacaaac	acatactaatt	ggaacattct	ttagaagagc	cccaaagtta	atgataaaca	1320
ttttcatcaa	agagaaaaaga	gaacaagggt	ttagcaaatt	cctctatcaa	ataacactaa	1380
acatcaagga	acatcaatgg	catgccatgt	ggaagaggaa	gtgctagctc	atgtacaaac	1440
cagtagataa	tttcaacttg	ctgccgaatg	aaacctcttt	gcaagggtatg	aatcagcact	1500
tctcatgttg	gttttgcttt	gtttttgttt	gttttttagag	acaggccctt	gctctgtcac	1560
acaggctgga	gtgcagtggc	acgatcagag	ctcactgcaa	cctgaaactc	ctgggctcaa	1620
gggatcctcc	tgcccttagcc	tcccaagtag	ctgggactac	aggccaccca	tgccagcta	1680
atcttttttaa	ttttctatag	agatgggagc	tcactagcac	ctttcatgtt	tgatgttcac	1740
atacaacgac	caagggtacaa	tgtggaaaag	ggtctcaggg	atctaaagtg	aaggaggacc	1800
agaaagaaaa	ggggttgcta	catagagtag	aagaagttgc	acttcatgcc	agtctacaac	1860
actgctgttt	tcctcagagc	agagttgatg	atctaaatca	ggggtcccca	acccccagtt	1920
catagcctgt	taggaaccgg	gccacacagc	aggaggtgag	caataggcaa	gcgagcatta	1980
ccacctgggc	ttcacctccc	gtcagatcag	tgatgtcatt	agattctcat	aggaccatga	2040
accctattgt	gaactgagca	tgcaagggat	gtaggttttc	cgctctttat	gagactctaa	2100
tgccggaaga	tctgtcactg	tcttccatca	ccctgagatg	ggaacatcta	gttgaggagaa	2160
aacaacctca	gggctcccat	tgattctata	ttacagtagg	ttgtatcatt	atttcattct	2220
atattacaat	gtaataataa	tagaaataaa	ggcacaaatag	gccaggcgtg	gtgggtcaca	2280

17/122

cctgtaatcc	cagcacttctg	ggaggccaag	gcaggcggat	cacgaggtca	ggagatcgag	2340
accatcctgg	ctaaaaacggt	gaaacccccgt	ctactaaaaa	ttcaaaaaaa	aatttagccgg	2400
gtgtgggtgg	gggcacctgt	agtcccagct	actcgagagg	ctgaggcagg	agaatggtgt	2460
gaacctggga	ggcagagctt	gaggtaagcc	gagatcacgc	cactgcactc	cagcctgggc	2520
gacagagcga	tactctgtct	caaaaaaaaa	aaaaaaaaaa	aaagaaataa	agtgaacaat	2580
aaatgtaatg	tggctgaatc	attccaaaac	aatcccccca	ccccagttca	cggaaaaatt	2640
ctcccacaaa	accagtccct	ggtgccaaaa	aggttgggga	ccgctaattc	aaataatcta	2700
atcttcattc	aatgctaata	aatgaataaa	ctttttttta	aatacacggg	ctcactttgt	2760
tgcccaggct	ggagtacggg	ggcatgatca	cagctcactg	tagcctcaat	cacccaggcc	2820
ccagcgatcc	tcccacctaa	acttcctgag	tagctgggac	tacaggcacg	caccaccatg	2880
cccagctaat	ttttaaattt	tttatagaga	tgggggtctc	accatggtgc	ccagactggt	2940
ctcaaaccct	gggctcaagt	gatcctccct	caaactcctg	gactcaagtg	atcctccttc	3000
cttggcctcc	caaagtgtctg	ggattacaag	catgagccac	tgtaccacgc	tggataaaca	3060
ttttaagtcg	cactacagtc	atggacaatc	aggcttttca	acatgcagta	tggacagtga	3120
gtcccagggg	ctgcttttcc	atactgaaat	acatgtgata	ctaaggagaa	agggtgctcgc	3180
aaggatattt	aaaatgaaga	atatttaaaa	tgaggaaaaa	actgtttctt	catgactttg	3240
ataaggctga	taaagacat	ttctgtgata	tcagggtgatt	cactcaagta	gtatatttca	3300
gtaatcattt	tctggaacag	cctgaatctt	aaccaaataa	ccatgatttt	ttaatgtctgt	3360
tatgatccct	tgatgatatg	accaaactgc	aatgtaggca	gctaaatctc	cacgagtttg	3420
actccccga	gagttgacag	ttttcttcac	aaattaaaga	aatatatttt	ttgatacatg	3480
attggcatat	ttaaaaacta	cactgaaatg	ctgcaaaaatg	atataaagaa	acatttttcca	3540
gaatcaaatg	caatcaaaga	gtggattagg	aatctactca	ccattatcaa	ctaaatagaa	3600
acacttggac	tgggtgtggg	ggctcacatc	tgtaatctca	gcactttggg	aggccaaggc	3660
agggtggaatg	cttgaggcca	ggagctcaag	accagcctga	gcaacatagc	aaaactctgt	3720
ctctacaaaa	aaaaaaaaaa	attaaccagg	catggtggga	gatgcttgta	atcccagcta	3780
ctctggaagc	tgaagtagga	ggactgcttg	agcccaggag	atcaagactg	cagtgaagcc	3840
tggtcatgct	gcgccacagc	ctgagtgaca	gagagagacc	ctgtctcaaa	aacaaaaaca	3900
aacaaaaaac	acttaacctt	cctgtttttt	gctgttggtg	ttgttggttg	tttgttttga	3960
gatggagctt	cactctgttg	cccaggctgg	agtgcagtg	cgtagctctg	gctcactgca	4020
agctctgcct	cccggtttca	cgccattctc	ctgectcagc	ctcccagta	gctgggacta	4080
taggcgcccc	ccaccacgcc	cggctacttt	tttgcatttt	tagtagagat	gggggtttcac	4140
cgtgttagcc	aggatggtct	tgatctcctg	acctcgtgat	ccactgcct	cggcctccca	4200
aagtgcctgg	attacaggca	tgagccaccg	caccgggcca	acctttctgt	tttttagttt	4260
gatatgcttg	ttaactcagc	agctgaaaag	atgctgaaag	tggccttcag	taaaaaaatt	4320
tcactagaat	ctctacatcc	atatttaatc	tgaatgcata	tccagattga	tcagttagag	4380
caaaaacact	catcatcatt	cctgatgacc	tctaattctg	gtttcggtct	tctatttcaa	4440
tggaaacaga	ataaggaaa	aaatggaagg	gctctggaaa	tttgtcctgg	gctatagata	4500
ctatcaaaga	tcaccaacaa	taagatctct	cctataaata	taaaaacaag	ataattaatt	4560
ttttaattat	ttttttctct	tcagaggatt	ttattttcaag	ataaaaacata	acttctaccc	4620
atactattga	ttccaaagg	tagaaaaagt	gtttttcctc	atcttatcct	tcaaagaggt	4680
cacagcaatg	caaacatcta	taaaatgcct	ctgcataatt	gtcagaagct	atagttccaga	4740
aatcattgaa	catctttttc	catttttaagc	ttaggtgagg	tgtcttagga	aacctctatg	4800
acaacttact	ctattttattg	ggagggtaaac	tcccagactc	tcccagggtc	tctgtatttg	4860
atctcatttt	ttaggcttcc	taatcccttg	aagcacaatc	gaaaaagccc	tggatctctt	4920
ttctgcacat	atcatcgccg	aattcattcg	gcttccagca	agctgacact	ccatgatata	4980
agcggcctcg	cccttctccg	gacgccagtc	cttgctgcgg	ttagctagga	tgaggggttt	5040
gctgggcttc	agtgcaggct	tctgcggtt	cccaagccgc	accagggtgg	ctcacaggct	5100
ggatgtcacc	attgcacact	gagctcctgg	caggctgtac	caatttttta	attattttaat	5160
atttattttt	aaaattatgg	tgaatatttt	ggtattctgc	tctaaaatag	gcccataaat	5220
gcacagcaga	tatctcttgg	aacccacagc	ctgaaaagtc	acatctctcc	tatttttctt	5280
ttaaagatgc	tactaagtc	gatctagcta	acagaaatcc	cagacacatc	attggttctt	5340
aagacttggg	atttatgaga	aagaggctca	gccctaagca	ggcccctgca	ccaggagggt	5400
cccagagtgc	agtcctccta	acttcccaag	gtgcatgctc	cagagaggca	ctgaaacagc	5460
gggtctgaga	cccacatagc	tggaagactc	aacggagggg	aacggagggg	gcccagtggtg	5520
tgagcacaag	cctgcaagcc	gaacacagag	aaatggatgg	caccaacaac	cagcaaagcc	5580
actaacataa	agagaaaagg	cagtgacggg	gccagaaccc	tcattcccaa	agactcttca	5640
ttcatggcca	atgaaagcat	tggctagaga	gtgaagttat	gatcatgtat	agagaggtaa	5700
ctgcctttag	tgaaaaaaca	actctgttaa	tgtgaaattc	cctatctgct	agactaaaag	5760
agttacattt	ttatattctg	tatccatta	gttgctagag	acttaaaatg	aacagaacgc	5820
tttcagacac	cctgttcaaa					5880

18/122

acattgtcag	gatgactatt	acaaaaaaat	caaaagacag	caagtattgg	tgaggatgta	5940
gagaaactgg	aacttttgtg	cactgtttat	gagaatgtaa	aatggagcag	ctgctgtgga	6000
aaagagtatg	caggttcctc	aaagagttaa	accaagatgt	ggaacaact	aaatgcccct	6060
cagtggatga	aggggtagac	aatatgtggt	atatacatat	catggagtac	tattcagcct	6120
ctaaaaaaa	aaaaggaaat	tctataacat	gcaacagcat	ggatgaatct	tgaggacatt	6180
ttgtctaatg	aataaggcag	tcatagaaag	acaaatactg	cacgactcca	cttatatgag	6240
ataccaaaaa	tagacaaatt	catagaatca	aagagtacaa	tggagggttac	ctggagctgc	6300
agggcgggaa	acgaggagtt	actaatcaac	gaacataacg	ttgcagttaa	gtaagatgaa	6360
taagctctca	agatcagctg	tacaacactg	tacctagagt	caacaataat	gtattgtaca	6420
cttataaatg	tgtaaagggt	agattaacaa	atgtagtaga	tccacaaatg	tggttaagtg	6480
ttcttaccac	agtaaaaata	aaaaagaata	tcaagcccag	gagttcgaga	ctagcctggg	6540
taacatgggtg	aaaccctgtc	tctacagaaa	atacaaaaat	tagccagctg	tggagggtgca	6600
ctcctagggg	ggctgagggtg	ggaggccttg	ttgagcccag	gaggtcaagg	ctgcagtgag	6660
ccatgaattg	accactgtac	tccagcccag	atgacagagc	aagacaccac	cccccccaaa	6720
aaaagaaaaa	gaatatcaaa	cattttaaaa	gatcagatac	gcaagaacaa	caacaaaaaa	6780
gagatgaaca	gagcatcgac	cctcatctag	tggtattctt	ggtctaactg	aaaaacagac	6840
attgagagac	aaacaatgac	agtgatgtga	tcacagcaat	tacacaggta	tcccctgggg	6900
actgcagaag	aaaggaggaa	tgccataactt	tcagaaaaata	gagaaagcgt	caaacagttg	6960
gtgaaagcct	tccaaaaacta	gagagaactg	cacacaccaa	atcacagaaa	gaagaaaagc	7020
cgtgggagat	tctgggaccc	accggctatt	tttgatggct	gaacaccctg	ctgcaggaga	7080
gacaggagct	ggaaagcatg	gtgggatgaa	acctcaaaac	gctttgcttg	cattgtctaa	7140
gatgactggg	cttgattaac	tctagtcaat	ggggacaatt	caatcaaaga	agaaagatgc	7200
tcaaattcac	attttagaat	gattttttat	ggcagtatgg	ggaatagatt	aaaagagagt	7260
gaagctggag	gcaagaaaact	tggttaaggag	caactgaaac	agtctagatg	ataaataata	7320
aactgacaga	gtgactagaa	aaatcagaac	aggctgaatc	aacagatacc	tagatgaaaa	7380
taacaggact	tgatcaccag	ttgtatcttg	gagaggaagg	agttgtttcc	ttgctttccc	7440
tacgactggg	aatacgggaag	gtttgcccgtg	tgtattgggt	atatactggt	gtgtagccaa	7500
tcaactgaaa	ccattttagca	gcttaaaaaca	caaaggctta	tctcccagtt	tctgtgggcc	7560
aggaatctaa	gataggctta	gctggctgggt	tctggctcag	agttttctca	gaggttgcaa	7620
tcaagatgtc	agctgggggt	gcatcatctg	aaggctcaac	tgggggccgga	gggtccactt	7680
ccaaggaggt	cactcacctg	cctgacaagg	cagtgtgggt	tggtggcagg	agatctcaat	7740
tcattgccaa	gtgagcctct	ctatagcatt	gctggaacat	cctccccatc	tggcagttgg	7800
cttctctcag	catgagtgat	ctgagagaga	gagcaaggag	gaagccacag	tggtcttctt	7860
actoctactc	ctaacactat	ggacctaactc	ctaacactct	cacttctgcc	ttattccatt	7920
agtttagaaag	ggaactaagc	tccacctctt	gaaataagaa	gtgtcaaaga	atthgtggat	7980
atatttataaa	atcatcacac	tgtggaagtg	gataggggggt	tcaattaatg	ctgaacttga	8040
aatgcctgag	acattcaaata	gtccaacagg	caatgaacat	acccatagat	gggtcatgact	8100
ttagcaagaa	tagaggaaga	tcacagaatt	aaggaggaat	tgaaggttaa	aagaagtggg	8160
gtcagattcc	ccctgaaaag	tgagccatga	aaggaaacttt	aactattgag	ttagagggtca	8220
gagttaggaaa	tttcggtgga	attctttttt	aaagaaagga	accatataag	catgttttga	8280
ggtagagggg	gaataaaatca	gtagacaggg	agaggttaaaa	aacataaatg	atagggggata	8340
gttgacaaaag	gtcttggcag	aatcccttac	ccattgactt	ggggccaaga	gagggacact	8400
tctttgtttg	agggataagg	aaaataagaa	agaatgggtg	ctatttagtg	tggtcctgtc	8460
tctagggcaa	acgcataagg	aacaaactgt	gtgtgttagg	aatatagatg	tgacctcaca	8520
ttgagattct	cacctcaaat	ccattttgtt	gttacctgta	ccttccctacc	ttctcttttt	8580
gctacatgca	gactgctgtt	ttgtcttctt	ggcctgttcc	aggtttcagc	attctggcat	8640
atctgtctacc	ctgttcccaa	acctctctag	agtcctatgct	ccttcccttg	atagtgtttg	8700
attggggccac	gtatctaaga	agtgtatgct	tcagtttaggc	ctgagaacct	cctctatgga	8760
aatctccatc	agtgaccctg	acagacttgg	tatcttggag	atgtcactgc	tcccagcctg	8820
tggtctagga	gaatctcagc	ctgggcctct	agtatgatgg	ataaggcgtt	aaggatatctt	8880
tgaaccagag	tctgtcatat	tcctcaatgt	gggacagata	aaacagtggt	agtgtgtgtg	8940
ttctgtagct	agaactctgg	tttttggtct	agattctttg	atgtatgacc	tttcagaggt	9000
attaaaaattt	gttctaatac	aatgttcaat	acaaatgtag	ttccttttct	gttaggacct	9060
caacaaaaaca	tgaccaactg	tagatgaaca	ttaaactatg	acaattcatg	gaaatgaata	9120
cagtaataacc	tgcggttccc	ccatttttagc	agtcactatg	gtgacatttg	gcacaaattg	9180
ctatttaagg	gtgcttttgt	taaaacctac	catcttacta	ggcacatgat	attgaaacta	9240
atgaaataat	ggagaaaact	cttaaaaaact	tttaatgaat	aaagtgatga	agtgataata	9300
ttttagctgc	tatttataaa	gtgactatta	caggtcaaac	attcttctag	gggttttttg	9360
ttgaagttgt	cacatttaata	ccttaataaac	ccactatgag	tcaggtattc	ttctctcccc	9420
tttggacagt	tggggaaaatg	gggggtcagag	aggttaggta	atthgtctcag	ggccacacaa	9480

19/122

cctgcatgta	gaaaaatctga	gatttgtaca	ggaacgtatc	aaactctgaa	gtccatgctt	9540
ctattttccc	atgctgcctt	tctaataaaaa	ggtaactaat	gctactggat	gctgccccca	9600
aagtgaagca	ctttcacccc	accctacttg	attttctcca	taaaactaat	cacatcctga	9660
caacttattt	attgctgata	tccccacta	gattataaac	tcaataaaaag	caagatcctt	9720
gtctgctgaa	tatcagtagc	taaaacgctg	tctagcacag	agcaagtaat	taatatattgt	9780
tgaatgaaca	aataaaggaa	aaaaattcaa	aggaagaaaa	agccctaaaa	cagatgttta	9840
cctaaacata	cattttaaaa	gaaagcatat	aacaaattca	ggacagaatt	taaatttgat	9900
tttttaaaaga	aataaccaag	tgctagctgg	gcacagtggc	tcacacctgt	aatcctagca	9960
ctctgggagg	ccgaggcagg	cagatcactt	gaggtcaaga	gttcaagacc	agcctggcca	10020
acatgggtgaa	acctgtctct	actaaaaata	cagaaattat	ccaggcatgg	tggcaggctcc	10080
ctgtaacccc	agctactcag	gaggctgagt	caggagaatt	gcttgaaccc	aggaggcaga	10140
ggttgcagtg	ggccaagatt	gcaccactgc	actccagcct	gagtaacaaa	gcaagactct	10200
gtctgaagga	gaaggaaaga	aagaaggaaa	gaaggaaaga	aggaaagaag	gaaagaagga	10260
aagaaagaaa	gaaagaaaga	aagaaagaaa	gaaagaaaga	aagaaagaaa	gaaagaaaga	10320
aagaaagaaa	aagaaagaaa	gaaagaaaga	accaagtgtc	tatttgggac	ctactatgct	10380
atgtttttcc	atgcacgcta	ttttcagtaa	agcagtttag	aaacttgcaa	gatcataaca	10440
acaaatatat	gcttctataa	ctctaaaatt	gtgctttaag	aagttcctct	ttaccagctc	10500
atgtatgcat	tagttttcta	agagttacta	gtaacttttt	ccctggagaa	tatccacagc	10560
cagtttattt	aaccaaagga	ggatgcttac	taacatgaag	ttatcaaagt	tgagcctaag	10620
ttgggccagt	tcatgttaat	atactccaga	acaaaaacca	tcctactgtc	ctctgacaat	10680
tttacctgaa	aattcatttt	ccacattacc	aaggagccag	ggtaggagaa	tatagaaaga	10740
ccaccaaga	atccttactt	ctttcagcaa	aatcaattca	aagtaggtaa	ctaaacacat	10800
gccctaacaa	tgaatagcag	attgtgctca	gaagaatgat	ctacaacatc	ttactgtgaa	10860
ggaactactg	aaatattcca	ataagacttc	tctccaaaat	gattttattg	aatttgcatt	10920
ttaaaaata	ttttaagcct	aaatttttaa	aggtttgata	ttggtacatg	aatagacaaa	10980
cagacatgga	ctagaccaag	aattaggttc	aaacatatat	aggaatttaa	tatacgataa	11040
atctagtatt	ccaaaggaac	caacaaatgg	tgttcagaca	gcaggatagg	catcaggaaa	11100
aacacagttg	ggcaccctac	cttactccta	acaccaggag	taactgaagg	agcaccaaat	11160
atattattat	ttttaattata	gttttaagtt	cagggtacg	tgtgcacaac	atgcaggttt	11220
attacatagg	tatacatgtg	ccatgttggg	gaggagcacc	aaatatttta	aagaaaaaaa	11280
ttggccaggg	gcggtggctc	acacctgtaa	tcccagcact	ttgggaggcc	aaggtgggca	11340
gatcacctga	ggctggggagt	tgcgagaccg	cctgagcaac	atggagaaac	cccatctcta	11400
ctaaaaatac	aaaattagcc	aggcatgggt	gcacatgcct	gtaatcccag	ctacttggga	11460
ggctgaggca	ggagaatagc	tttaattctg	gaggcacagg	ttgcggtgag	ctgagatatt	11520
gcactccagc	ctgggcaaca	agagcaaaac	ttcaactcaa	aaaaattaat	aaataaaata	11580
aaataaagaa	agaaaagaaa	aaaatgaaaa	tagtataatt	agcagaagaa	aacaccgtag	11640
aatcctcgga	ctcttaggat	ggggaatgcc	tataatataa	aaaccctgaa	gttataaaaag	11700
agaaaatcac	ctacatacaa	accaaattct	tctacatgcc	taaaacatag	cacaaacaca	11760
gctaaataat	catagctgaa	tgaactggga	aaacaaaact	tgactcatat	ccagacagag	11820
ttaattttcc	tacacataaa	gagtacctat	ataaacccaa	caaaaaaacc	accactaacc	11880
caaaaataaaa	atgtgacagg	taatgaacag	gtagttcaca	gagaatacaa	atggctcttc	11940
ggcacataag	atgctcagac	tgacttttac	ttatttattt	tttgagagac	aggggtctcac	12000
gatgttgccc	agggttaggct	caaactcctg	ggctcaaatg	atagtaccag	gactacagg	12060
gtgccccacc	gcacctgggt	cctcaaccac	ctgtattaac	aggaaatgca	aaataaaaact	12120
ttcaaatcta	ttttacctat	tagaatggca	aaaatttgaa	aaacttcaaa	catcatcatg	12180
ttggtgagaa	tgtagggaga	ctggcactct	cattttttgc	tgatagcata	tataactga	12240
tggcttctat	ggaaagcaat	ctggcagcgt	ctatcaaatg	tacaagtgca	tatatccttt	12300
gacaaagcaa	ttccactcta	ggaatgtgtt	ctatatgggt	gtgcttcctg	gggctgggaa	12360
ctgggagcta	agggacaggg	gcagaagata	atcttctttt	ccctccttcc	ccgttaaaca	12420
tggtgaattt	tataactgtt	aatatattct	ttttcacaaa	agataatttt	taagcgatat	12480
gtctgggaat	tttttttttt	cttttctgag	acagggtctc	actctgtcat	ccaggctgga	12540
atgccatggt	atgatctcag	ctgactgcag	cctcgacctc	ctgggttcaa	gcaatcctcc	12600
cacctcagcc	tcctgagtag	ctgggactac	aggcacgtgc	catcatgcta	atttttgtat	12660
atacagggtc	tcactatggt	gcccaggcta	atgtcaaaat	cctaggctca	agcaatccac	12720
ccacctcagg	ctccaaagtg	ctgggattac	atggctgagc	caccgcgcct	ggccctggga	12780
attcttacaa	aagaaaaaat	atctactctc	cccttctatt	aaagtcaaaa	cagagaagga	12840
aattcaacct	ataatgaaag	tagagaaggg	cctcaaccct	gagcaacaaa	cacaaaggct	12900
atttctgaga	caggaatttg	ctgaacaaaa	tgcagggaag	atgacaagaa	tcaagactca	12960
cttctcgggt	gggcgcagtg	gctcacacct	gtaatcccag	cactttggga	ggccgaggcg	13020
gacagatcac	gaggtcagga	gattgagacc	atactggcta	acacagtga	acccagtctc	13080

20/122

tactaaaaat	acaaaaaatt	agccggggcgt	ggtgggcaggt	gcctgtagtc	ccagctactt	13140
gggaagctga	ggcaggagaa	tggcgtgaac	ccaggaagcg	gagcttgacg	tgagccgaga	13200
tcacgccact	gcactccagc	ctgggtgaca	gagcaagact	ctgtctcaaa	aaaaaaaaaa	13260
aagactcatt	tctctagatc	ttgagccgta	ttcaaattta	tctcagctta	gtgagagggt	13320
aaagcaagga	atatccttcc	ctgtggggccc	tgctccttac	tgaaggaagg	taacggatga	13380
gtcaaggaca	ccaatggaga	aaagcactaa	caccattatc	tgatgaacat	tacgtgaaga	13440
agggtaagaa	gtgaagtgga	attgctgaag	aagtacgtga	aagcggacat	tcatttgggg	13500
aaatgggaata	taggaaatcc	ataaaaagtga	ttaaaaagat	gttagaggct	gagggcggggg	13560
gaccacaggg	tcaggagatc	gagaccatcc	tggtctaacac	ggtgaaaccc	catctctact	13620
aaaaatacaa	aaaatttagcc	aggcgtgggtg	gcaggcacct	gtagtcccaa	ctactcggga	13680
gactgaggca	ggagaatggc	atgaacctgg	gagacggagc	ttgcagttag	ccgagatcac	13740
gccactgcac	tccagcctgg	gtgacagagt	gagactccat	ctcaaaaaaa	aaagttagat	13800
acgagagata	aagatccaac	agacacacaa	ctgctaattc	tgaacagaac	aaaacaaatg	13860
gcacagggaaa	agaaaaattta	agatataaca	ccggaaaact	ttcctgaaat	tgagtaactg	13920
aatctatagc	ttgaaagggt	ttagcatatg	ccaagaaaaa	tcagttagagt	ccaaccagca	13980
caagacacat	ctagcaaggc	tggtgattct	accaacacag	agaaagaagt	gggtgaccca	14040
taatgcggaa	aaaggcagac	catctgcagt	cttctccaga	acactggagt	ctgaagacaa	14100
aagaatgctg	cctactgagc	cagaaggagg	agaaagtga	ccaacacatc	tttaccagt	14160
tagaatgtca	cgcattatct	aaaggctgca	aaagccatga	aagacatgaa	agaacacaag	14220
catttacaac	atgaaagaac	acaagcattc	tcatactcaa	gaatccttaa	gaaaaatgta	14280
gtcctaatac	agcccactga	aagttaaatg	tacttaattg	gtcattaat	gggaacttca	14340
tagcttcaaa	tcagtctggg	cccatctacc	aacatctctc	gcccggcttt	cctgcaatag	14400
tcagcacctt	tccctcctcc	cagtcttgct	ccctggagtc	tgctctcagc	atagcagagt	14460
gaccacatca	acacccaagt	cagagccctc	cagtgcgcac	tggtctacaa	agcccttccc	14520
acccccacc	ccacgtgccc	tccggtacct	tgtgacgtgt	ctcctgcata	ccctagcagc	14580
cctggcctcc	tccactgcccc	tccgtgacat	cagggaaggc	actccttgag	tcttggctct	14640
ggcgcgctcc	tccacctgca	gtgagttaac	tcccttacct	actctaggct	attgctcaaa	14700
tgtcagcatc	tcaatggggc	cctccctgac	taccctatct	aaattctaca	tactcccctt	14760
gaccccatgg	acctcactca	ccctattcca	cttttattct	tacaatttag	cacttggtct	14820
cttctaacgt	attctaagac	ttactcattt	attacattgt	ttgccacccc	ctctagtaca	14880
taaactccag	aggggcaggg	atctctgtct	atcttattcat	ttctttatcc	ctaggacata	14940
gaacagggca	tagttcagag	tattcaatgt	tatcaatgaa	tgaactagca	gtagtaccag	15000
ttccagttag	gcacagaatt	aaatctaatt	agaattaaat	ctcatgggtc	gggttaacta	15060
tggatagaaa	attagatata	atcttaagaa	gcctagaaa	aaaaaattaa	taatgtaaaa	15120
ataatattaa	tttgataata	ataacaaaaa	ctctgccagg	cactgtgggt	caaatctgca	15180
atcccagcta	ctcaggaggc	tgaggtggaa	ggatcacttg	agaccagagt	tcaagactca	15240
gcctaggcaa	cacggcaaga	aactgtctct	aaaaaaatta	aaacttaaat	ttttaaaaaa	15300
gaattctcaa	agcgtcacaa	ttaaactgaga	taaggtaca	ggaagtgtga	agtaatatata	15360
ctatgctaata	gggttttttt	tttttttagaa	aggtataacc	aaaagatttc	ttctctcaagt	15420
cgataaactg	agaaaagataa	gcatatcttc	caattaacag	aggggggagga	aaagccagat	15480
acaacaaaat	aagatataaa	ttagtttcca	gttgaaaaaca	agagttaggag	ttattttgca	15540
tcacctcacc	tgtgacctcc	cccagcccaa	aaaacactac	tgataaacag	ggtagaaaag	15600
catcatctca	gataaagcag	gaaaaactgc	cacagtctca	aaccacaaac	tataagcaca	15660
cacctggcca	accctgcca	gtctgggctc	agtaggagga	acgtgctgag	agctaggatg	15720
taccaactta	gacattctgt	gggatacaga	tgtccctgga	agggtcacac	catctcaaag	15780
gcacctgtaa	tgcccactga	ttacagccac	catatgtgag	agagaaactc	agggcactta	15840
gagagtataa	caagaacctt	atgtcatctg	agatgaggaa	tcctcagccc	tgcaaattaa	15900
ccaactcttt	agaacaactg	gcaaaacata	aatatccaca	acttttgggt	cagtaattcc	15960
actcttagat	atcaatccaa	agtagatgag	acagcagata	cacacacaaa	atgggtattta	16020
ctgcagcatt	gtttataata	gcaaaaaaca	agaaataatc	catatgtctc	aataggatag	16080
tgggtacatg	aggggtatgta	cccatcattc	aacctcaaaa	aagagtgata	tggtgttcca	16140
cagatggaca	taaaaagctg	tgtgttacgt	gaaaaacaa	tcaagcagca	gcaggatggg	16200
cttatgatag	tcagtatgag	ctaatttctg	gaaaaaaaaa	tctagtgtgt	gcacagaaaa	16260
catctgaaag	aacagaaaca	aaactatcag	cagaatattg	agatgtttta	ctaagtgtga	16320
tatctatact	gcttgaattt	tttaccctaa	gcaagaatta	ctttttggaa	aaagaaaatt	16380
caggaaataa	agcattttctt	ttaaacttcat	gttttaacaa	atgggtgatg	aataaaaagag	16440
ttcttattca	tcataaacac	acacagcaca	catgcacgca	tgtgcgtgag	cacacccttt	16500
acttgataaa	taccatgttg	aatattttag	tctttccttt	taggttctat	cccttcactc	16560
aaaatgcggg	tataaataaa	tgtacttttc	atgtgccttc	tgcctaaacc	cactttaata	16620
taactttaca	gtcccattat	cattatagtc	tcaaagctag	actcagcctg	aaactaccct	16680

21/122

ttcatttggga	acccttatta	aaatgccaca	tacagctcct	tcaaataaaa	acaaacccta	16740
ggccctgaca	ctaggettcc	tttgttgcta	ctcataatgg	ccaagttctg	tgcttataat	16800
acatcttctt	tcattttatt	gctacatata	caagggtttt	atatgttttt	cttattatat	16860
cttaattcaa	aacaccatca	cgctcttttc	cagatgaaaa	taaggaaaaag	aaatttgagca	16920
actgactgac	ttaaaggtca	taaaactata	tagtagcaga	gtcagcaaaa	gaagaaacac	16980
acatctccca	agtagaggct	gaaaaccagt	accattcacc	tccaggggtga	gctatataca	17040
gattacaaag	tcacctttct	taaatgttca	aactgaatcc	catacccata	ctttaccact	17100
acctcgtaag	aacagcctca	gatcttggtta	tagccttttt	tttagcatgc	tgaagccaat	17160
aaaatgcttc	ccatttcagca	agagaaacaa	gttctgaaac	actgaataat	ctgcccaggg	17220
cctatgaaca	tttccactgt	gagaaaatgt	ctccactgtg	tggaagaagat	ccttactctt	17280
ctccacacag	gcagaacatt	agaaaaattc	ttggattcta	tgatgcacag	cttaggagtc	17340
tgtttagcag	aattttaagtc	caaatagtta	ttaaatcctc	ctctgttcca	gaaacagtgc	17400
taataactgt	gaatataaaa	attgaaaaga	tactctcctg	gctcccaaga	aagtcagcca	17460
gataatggag	acacaggcac	acaaatcact	gtcacatgaa	gctctacctc	cctaacttca	17520
aaatggggcc	taagtccacca	agaatacagt	agcagttgtg	actacgagta	actactataa	17580
ttcatttctt	tatcttccct	tagaaaactc	ttctcccttg	gaaatttatt	tgcatcttcta	17640
aatagccttc	cttactaaaa	ggaagcaggg	ctccttgggg	aaatagctga	ttctaggtgt	17700
ggactatgan	atgaaaatgg	tgagtctggg	acatcccctg	ttgcccagaa	atcaaggaac	17760
tgcccaatga	ttaacagagt	catgtttaa	ggacctaaga	gtgaaccaga	aggagctcac	17820
tttggggggc	gtggaacaat	ttcaagaaaa	acatgacagt	aatgaattat	aaaacatgaa	17880
ttaaatataa	tattggtact	aaaaagagaa	caaaaggatg	tggtcttggga	taaagctctt	17940
cttcatgaaa	gaataaccagc	aaaggaaatg	aaaggaaatg	agagaattag	aaaaattatc	18000
attttttaaa	ccttaatata	ttcacctaga	catgctaaaa	ccactgagta	aaaggctgct	18060
tggttatagg	atgctcacat	gatctcagag	tttcacacca	cagataattt	attagataca	18120
ggagttatga	tgtgatcaag	cttctctgtg	ccccacagcca	ggccccacaa	cactatgtgc	18180
ctccttttga	tgtgggagct	acacagcatc	gcccacacag	cttctcgcca	aaactgtttg	18240
aagcttatca	caaggggaaga	actggacagc	ttctgacctt	gagacgctcc	accagacaac	18300
ttgcttggcc	tctccaaaga	aacttgcttg	gcctctccaa	agaaaactca	gtttcattta	18360
aaaacuaaac	taattattta	aaaacaaacg	aaaagcaagt	tgtggacttg	agctccaggg	18420
acagagcaga	catacttttc	cctgttcttc	ccagtaagtg	gtaataaaaa	ccctcaaac	18480
tagatataaa	acaaatataa	gaaggttctg	gaaggggaag	aggaggcaga	ctatccaggt	18540
gccttggggc	ccacagaaca	acccagtgat	gggttctactg	ggtcttcttt	ttgcttcttt	18600
atctcagact	tggaagctgaa	gcagcaggca	acttcaaaaac	accaagggggc	acagattgaa	18660
aagccccaag	aaaagcctgc	cctctctagc	caaaggacca	ggaaggagac	agtctaataa	18720
gatggaacac	atttagacag	taactgccca	tttaccagca	ataactgagc	agggagccta	18780
gacttccagt	cttgtgagga	cgtaccaagg	taccaaacac	ccccaccaag	gctgagtaag	18840
gactgcgact	tttatccctg	catggcagta	gtaaggagcc	catccctcac	ccgccagcag	18900
tgtcagggga	acctggactt	ccactcccat	ccaggagtga	tgaggccctc	cctgctgggg	18960
tcattgtcaga	ggaggcctag	tggaatttca	gtgacttaac	cttttcccag	agataatgag	19020
gccacctttc	ctccctcttc	ccccatggtg	acagtgaag	cactgtggca	agcagtaggc	19080
actcctaccc	ctcctagcca	gggagggtatc	agggaggcca	agtagggaac	cagaataccc	19140
acaaccaccc	agcagcaaca	gggggtcccc	accccatgtg	gtgtcaatgg	aagcagagcg	19200
gaaagcctgg	atattttacc	ccatctagaa	gtaacaagct	gatgtcccc	ttcttctact	19260
acaatggtgt	tcaaaacagg	tttaaaataa	gtctagagtc	tgataacgta	ataccctaat	19320
cgttgaagtt	ttcatttgagg	atcattttata	ccaagagtca	ggaagatccc	aaactgaaag	19380
agagaaaaga	caattgacag	acactagcac	taagagagca	cagatattag	aactacctga	19440
aaggatgtta	aagcacatat	cataagcctc	aacaggctgg	gcgcgggtggc	tcacgcctgt	19500
aacccccagca	ctttggggagg	ccgaggcagg	tggtacacaa	gatcaggaga	tcgagaccat	19560
cctgggtaac	acggtgaaac	cccgtctcta	ctaaaaatac	aaaaaaaaat	agcaaggcat	19620
ggtggtgggc	acctgtagtc	ccagctactc	gggagcctga	ggcaggagaa	tggtcatgaac	19680
ctgggaagag	gagcagtgag	ccgagatcgc	accaccgcac	tccagcctgg	gcaacagagc	19740
aagacttcgt	cccaaaaaaa	aaaaaaaaaa	aaaaaaaaagc	ctcaacaaac	aactacaaac	19800
gtgcttgaaa	caaataaaaa	aaaaatcttg	gcaaagaaat	aaaagatata	tattttggcc	19860
aggtgcagtg	gctcacagcc	tgtaatccct	gcactttggg	aggctgaggc	aggcggtatca	19920
cctgaggtca	ggagtttgag	accagcctga	ccaacatgga	gaaacccctg	ctctactaaa	19980
aatacaaaaat	tagccagtc	tggtggcaca	ctcaggaggt	cctagctact	caggaggccg	20040
aggcaggaga	atcgcttgaa	ctcaggaggt	ggaggttgcg	gtgagccgag	atcccgccat	20100
tgacacattgc	actccagcct	gggcaacaag	agcaaaaactc	catctcaaaa	aaatagatac	20160
atatttttaat	ggaaatttta	gaattgaaaa	atacagtaac	caaattgaat	ggaaagacaa	20220
catagaatgg	agggggcaga	caaaataatc	agtgaacttc	aacagaaaaat	aatagaaatt	20280

22/122

acccaatatg	aagaacagaa	agaaaataga	ctggccaaaa	aataaagaag	aaaaaagagg	20340
agcagcagga	ggaatgatgg	aaaaagagaa	aggaaggaag	gaagggaagg	agggaggggaa	20400
ggagtggagg	agaaagtctc	aaagacctct	gagactaaaa	taaaagatct	aacacttgtc	20460
atcaggggtcc	aggaagagaa	caaagatggc	acagctggaa	acgtattcaa	aaaataatag	20520
ctgaaaactt	cccaaatttg	gcaagagaca	taaacctata	gattcgaaat	gctgaacccc	20580
aaataaaaaag	cccaataaaaa	tccacaccaa	aatacatcat	agtcaaactt	ctgaaaagac	20640
gaaaagagaa	aacgtcttga	aagcagttag	tgaacaaca	cttcatgtat	aagggaaaaa	20700
caattcaagt	aacagatttc	ttacagaaat	taagggaagcc	agaaggaaat	gacacaatgg	20760
ttttcaagtg	ctgaaagaaa	agaagtgtca	acacaaaatt	ctagattcag	taaaaatatc	20820
cttcaagaat	caatgggaaa	tcaagacagt	ctcagataaa	gcaaaaataag	agaatatgtt	20880
gccagcagat	ctcccctaaa	ggaatggcaa	aagggaagatc	atgcaacaga	ccaaaaaatg	20940
atgaaagaag	gaatccagaa	acatcaagaa	gaaagaaata	acatagtaag	caaaaaatata	21000
tgtaattaca	ataaaaatttc	tatctcctct	taagacttct	aaattatatt	gatgggtgaa	21060
gcaaaaattta	taaccctgtc	tgaagtgtct	ctactaaatg	tatgcagaga	attataaatg	21120
gggaaagtat	aggttttctat	acctcattga	agtggtaaaa	tgacaacact	gtgaaaagtt	21180
acatacacac	acacagtaa	gtatatataa	atatatgtgt	gtatatgtgt	gtgtatatat	21240
atatatacat	ataatgtaat	acagcaacca	ctaacaacac	tatacaaaga	gataataacc	21300
aaaaacaatt	tagataaatt	gaaatggaat	tctaaaaaat	attcaaatac	tctacaggaa	21360
gacaagacaa	aaagagaaaa	aaagaggagg	acaaactaaa	ttttttaaaa	acataaataa	21420
aatggtagac	ttaagcccta	acttatcaat	aattacataa	atgtaaatga	tctaattata	21480
tcaattaaaa	gacagagata	gcagagttaa	tttaaaaaa	tagctataag	aaacctgtct	21540
tgggctgagt	gcagtgactc	acacttgtaa	tcccagcact	tcgggaggcc	aaggcgggtg	21600
gatcacctga	ggtcaggagt	tccagaccag	cctggacaac	atggtaatac	cccatctcta	21660
ctaaaaatac	aaaaaaaatta	gccaggcatg	gtggcacacg	cctgtagtcc	caactactca	21720
ggaggctgcg	acacaagaac	tgcttgaacc	cgggcagcag	aggtagcagt	gggccaagat	21780
tgccgcactc	cagcctgaac	gacagagtga	gactccacct	cagttgaaaa	acaaaaaaga	21840
aacctgtctt	aaatatacca	acatatgttg	gttgaaatta	aaagaataaa	atataatcatg	21900
aaaacattaa	tcaaaagaaa	ggagtggcta	tattaataac	ataaaataga	cttcagagaa	21960
aagaaaattt	caagagacag	gaataaaaag	atcaagaaaa	gatcctgaaa	gaaaagcagg	22020
caaatcaatc	attctgtctg	gagattcaac	accctctctt	aacaactgat	agaacaacta	22080
gacaaaaaaa	tcagcatgga	gttgagaaga	acttaacacc	actgaacaac	aggatctaata	22140
agacattttac	ggaacactct	acccaacaat	agcaaaaata	acattctttt	caagtattca	22200
ctgaacatat	ccttagacc	taccctgggc	cataaaacaa	agctcactag	tgattgccga	22260
aggcttggat	ggacagtgga	agagctgcac	ggggaggag	aaggtgacag	ttaaagagtg	22320
taggattttct	ttttgggata	atgaaaatgt	tccaaaattg	attgtgggtga	tgttggcgga	22380
actctacaaa	tataaaaaag	gccattgaat	tgtacgtttt	aagtgggtga	aacatatggt	22440
atgtggatta	tatctaagc	tttttaaaaa	cttaacacat	ttcaagaat	agaagtcata	22500
cagagtgtgc	tctactggaa	tcaaaactaga	aagaggtaac	tggaggataa	cgagaaaaagc	22560
ctccaaatac	ttgaaaactg	gacagcacat	ttctaaaatc	atccgtgggt	caaagatat	22620
catttctgat	attcattttt	attgtttaat	gtatttttaa	aaatttctta	agggaaataa	22680
actgactaaa	aatgaatatg	gctgggtgcg	gtggctcacg	cctgtgatcc	cagcactttg	22740
ggaggccgag	gctgggtgat	cacaagatca	ggagttcgag	accagcctgg	ccaagatggt	22800
gaaaccccg	ctcaactaaa	aaactacaaa	aagtagccaa	gcgcagtggc	gggagcctgt	22860
ggtcccagct	acttgggagg	ctgaggtagg	agaatcgctt	gaacacaggc	agcagaggtt	22920
gcagtgaagc	aagattgtgc	cactgcacgc	cagcctgggc	gacagagact	gcctcaaaaa	22980
aaaaaaaaaa	aaaaaagaata	tcaaaatttg	tgggacatag	ttaaagcaat	gctgagaggg	23040
aaatttataa	cactaaatgt	ttacattaga	aaagagaaaa	agtttcaaat	caatagtctc	23100
cactcccatc	tcaagaacac	agaagatgaa	gagcaaaata	aacccaaagc	aagcaaaaga	23160
aagaaaaaat	aaaaataaat	cagtaaaatt	gaaaacagaa	acacaataaa	gaaaatcagt	23220
gaaacaaagt	actgattctt	cgaaagatta	ataaaattga	caaacctcta	gcaaggctaa	23280
caaacaaaaa	agaaagaaga	cacggattac	cagttattag	aatgaaagca	taattagaaa	23340
caactctaca	cattataaat	ttgacaatgt	agatgaaatg	gactaattac	tgaaaaaaca	23400
caaattacca	caactcacc	aatatgaaat	agataattgg	gatagcctga	taactactga	23460
gaaaattgaa	tttgaattt	taacactctt	aaaacagaaa	cattaaaactt	aatattttat	23520
aaatattaga	taaggttaatt	atacccttcc	ttacaacata	aaaacgacaa	attattttgc	23580
agctaaagag	atgtatgtac	tgtgaaaaat	atcttcagaa	aaatagaact	ttgtttgaag	23640
aataaggatt	taaaaaatgt	ttttaactct	caagaagcaa	atatctgggc	ccagatgggt	23700
tactgaaga	attctaccaa	atgtttaatg	aagaattacc	accaactcta	catagcatct	23760
ttgagaaaac	tgaagagaag	ggaacatctc	ccagttcatt	ttatgaagtg	ggtgttactc	23820
tgataactaga	actgtataag	gacagctact	cttgacacac	tgccatggg	tagctctgct	23880

23/122

ctgcaggaac	agtcagaaaa	aaaaaaaaaa	gaagcactgg	acaagggcag	tataaaaaaa	23940
gaaaactggg	ccaggtgcag	tggctcacac	ctgtaatctc	agcacttttg	gaggctgacg	24000
ctgggtggatc	acctgagggtc	aggagtgtga	gactagcctg	gccaacatgg	taaaaccctg	24060
tctctactaa	aatacaaaaa	ttagccaggc	aggggtgggtg	ggaaaaataaa	aaggaaaaaa	24120
aaacaaaaat	aaactgcaga	ccaatatcct	tcatgagtat	agacacaaaa	ctccttaaac	24180
tccttaacaa	aatatttagca	agtagaagca	atatataaaa	ataattatac	accatgatca	24240
agtgggactt	attccagaaa	cgcaagtctg	gttcaacatt	tgaaaaacaag	gtaaccctact	24300
atatgaacgt	actaaaggag	aaaactacat	aatcacatca	atcaatgcag	aaaaaagcat	24360
ttgccaaaat	ccaatatcca	ttcatgatac	tctaataaga	aaaataagaa	taaaggggaa	24420
attccttgac	ttgataaagc	ttacaaaaga	ctacaaaagc	ttacagctaa	cctatactta	24480
atggtgaaaa	actaaatgct	ttcccctacg	atcaggaaca	aagcaaggat	gttcaactctc	24540
attgctctta	tttaacatag	ccctgaagtt	ctaacttggtg	caaaacgata	agaaagggaa	24600
atgaaagacc	tgcagattgg	caaagaagaa	ataaaaactgt	tcctgtttgc	agatgacatg	24660
attgtctcat	agaaaatgta	aagcaactag	gggtaggggg	gcagtggaga	cacgctgggtc	24720
aaaggatacc	aaatttcagt	taggaggagt	aagttcaaga	tacctattgc	acaacatggt	24780
aactatactt	aatatattgt	attcttgaaa	gtactaaaag	agtgggtggt	aagcgtttctc	24840
accacaaaaa	tgataactat	gtgaagtaat	gcatacgtta	attagcacia	cgtatattac	24900
tccaaaacat	catgtttgtac	atgataaata	cacacaattt	tatctgtcag	tttaaaaaa	24960
catgattttg	gccaggcaca	gtgggtcata	cctgtaatcc	cagcatttta	ggaggctgag	25020
gcgagcagaa	aacttgaggt	cgggagtttg	agaccagaat	ggtcaacata	gtgaaatccc	25080
gtctccacta	ataatacaaa	aattagcagg	atgtgtgtgg	gtgcacctgt	agaccagct	25140
acttggggagg	ctgaggcacg	agaattgctt	gaacaagggg	ggcagaggtt	gcagtggact	25200
gggtgccact	gcattccagc	ctggtgacag	agtgcagactc	catctcaaaa	aaaataaaat	25260
aaagcatgac	ttttcttaaa	tgcaaaagcag	ccaagcgcag	tggctcatgc	ctgtaatccc	25320
accactttgg	gaggccgagg	caggcagatc	acaaggtcag	gagtttgaga	ccagcctgac	25380
caacatggtg	aaacccccatc	tctactaaaa	aatatataaa	ttagccaggc	atgtgtagtc	25440
tcagctactc	aggaggctga	ggcaggagaa	tcacttgaac	ccggaggcag	agggtgcagt	25500
gttgagccac	cgcactccag	cctgggtgag	agaacgagac	tccgtctcaa	aaaaaaaaaag	25560
caaaaataacc	taatttttaaa	aacactaaaa	ctactaagtg	aattcagtaa	gtcttttagga	25620
ttcaggatat	atgatgaaca	tacaaaaatc	aattgagctg	gacaaaaggag	gattgttttta	25680
ggtcagtagt	ttgaggctgt	aatgcacaat	gattgtgcct	gtgaatagct	gctgtgctcc	25740
agcctgagca	gcataatgag	accacatctc	tatttaaaaa	aaaaaaaaaatt	gtatctctat	25800
gtactagcaa	taagcacatg	ggtactaaaa	ttaaaaacat	aataaaatact	gttttttaatt	25860
gcctgaaaaa	aatgaaatac	ttacatatata	atctaacaata	atgtgcagga	cttgtgtgct	25920
gaaaactaca	aaacgctgat	aaaagaaatc	aaagaagact	taaatagcgt	gaaatatacc	25980
atgcttatag	gttggaaaaa	ttaatatagt	aaagatgcca	attttatcca	aattattaca	26040
caggataaca	ttattactac	caaaaatcca	gaaaaatgtt	acatagatat	agacaagatc	26100
atacaaaaaat	gtatacggaa	atatgcaaaag	gaactagagt	agctaaaaca	aatttgaaaa	26160
agaaaaataaa	agtgggaaga	atcagtctat	ccagtttcaa	gacttacata	gctacagtaa	26220
tcaagactgt	gatattgaca	gagggcagac	tatagatcaa	tgcaaccaaa	tagagaacta	26280
agaaagaagc	acacacaaat	atgcccacaa	gatttctgac	aaaggtgtta	aaacacttca	26340
acgggggaag	atatgtctct	cattaaaggg	tgtagagtca	ttgcacatct	ataggcaaaa	26400
agatgaacct	gaacctcaca	ccctacagaa	aaattaactc	aaaatgactc	aaggactaaa	26460
cataagatat	acatctataa	aacatttaga	aaaaggccac	gcacggtggc	tcacgctcgt	26520
aatcccagca	ctttgggagg	ccaaggcagg	tggatcacct	aaggtcagga	gtttgagacc	26580
agccggatca	acatggagaa	gccccatctc	tactaaaaat	acaaaattag	ctggacgtgg	26640
tggcacatgc	ctgtaatccc	agctacttgg	gaggctgagg	catgagaatc	gcttgaacct	26700
ggggggcaga	ggttgcggtg	agccaagatc	acaccattgc	actccagcct	gggcaacaag	26760
agcaaaaactc	caactcaaaa	aaaaaaaaaa	aaaggaaaaa	tagaaaatct	ttgggatgta	26820
aggcgaggta	aagaattctt	acacttgatg	ccaaactaag	atctataaag	ccagtcgtgg	26880
tggctcatgc	ctgtaattcc	agcactttgg	tcaactagat	gaaaggtata	tgggaattca	26940
ctgtattatt	ctttcaactt	ttctgtaggt	ttgacatttt	tttagtaaaa	aattggggga	27000
aagacctgac	gcagtggctc	acacctgtaa	tcccagcact	ttgggaggcc	ggggcagggtg	27060
gatcacacgg	tcaggagttc	gagaccagcc	tggccaacat	ggtgaaaccc	cgtctctacc	27120
aaaaatataa	aaaattagcc	gggtgtcatg	gtgcattgct	gtaatcccag	ctactgagga	27180
ggctgaggca	ggagaatcac	ttgaacctgg	gaggtggaag	ttgcagtgag	ccgagattgt	27240
gccactgcac	tccagccttg	ggtgacagag	cgagactccg	tctcaaaaga	aaaaaaaaaa	27300
aaagaatatc	aaacgcttat	tttagaaaat	atttaaagga	gccagaattt	aattgtatta	27360
gtatttagag	caatttttat	gtccatggc	attgttaaat	agagcaacca	gctaacaatt	27420
agtggagttc	aacagctggt	aaatttgcta	actgttttagg	aagagagccc	tatcaatata	27480

24/122

actgtcattt	gaggetgaca	ataagcacac	ccaaagctgt	acctccttga	ggagcaacat	27540
aaggggttta	accctgttag	ggtgttaatg	gtttggatat	ggtttgtttg	gccccaccga	27600
gtctcatgtt	gaaatttgtt	ccccagtact	ggaggtgggg	ccttatttga	aggtgtctga	27660
gtcatggggg	tggcatatcc	ctcctgaatg	gtttgggtgcc	attccttgca	gaatgagtga	27720
gttcttcttc	ttagttccca	caacaactgg	ttattaaaaa	cagcctggca	ctttccccc	27780
tctctcgttt	cctctctcac	catgtgatct	cactggttcc	ccttcccttt	atgcaatgag	27840
tgggaagcagc	ctgaagccct	cgccagaagc	agatagtgat	gccatgcttc	ttgtacagcc	27900
tacaaaacca	tgagcccaat	aaaccttttt	tctttataaa	ttatccagcc	tcaggtattc	27960
ctttatagca	agacaaaatga	accaagacag	ggggaaatca	acttcattaa	aataatctat	28020
gcagtcacta	aacaaataag	aacaagaggg	tccagaagtg	ggaagccaat	acccagaggt	28080
cctacaatac	agtatctgaa	aagtccagtt	tccaaccaa	aaatatatat	atacaggccg	28140
gacatggtag	cttatgtctg	taatcccagc	actttgggat	gctgaggcgg	gcagatcacc	28200
ctaggtcagg	agtcgagac	cagcctggcc	aatatggcaa	aaccccgctc	ctactaaaaa	28260
tacaaaatt	agccaggcat	ggtgggtggg	gcctgtaatc	ccagctactc	gggagggtga	28320
ggcagggaat	cacttgaacc	caggaggcag	aggttgagct	gagccgagat	cagccagctg	28380
aactccagcc	tgggcaacaa	agtggagactc	cacctcaaaa	aaaaaaaaaa	tatacatata	28440
tatatgtgtg	tgtgtgtgtg	tgcgcgcgtg	tgtgtatatata	cacatacaca	tatatacata	28500
tatacagaca	cacatatata	tatgaagcat	gaaagaaac	aaggaagtat	gaaccatact	28560
ttctcgtgtt	atgataggat	ggggtatcac	gggggaagta	gacaagggaa	actgcaagtg	28620
agugtaata	gttatcagat	ttaacagaaa	aagacttttg	agtaaccatt	ataaatatgt	28680
ccacaaatt	aaagaaaagc	gtgattaaaa	aaggaaagga	aagtatcata	acaatattac	28740
tccaaataga	gaatatcaat	aaaggcatag	aaattataaa	atataatata	atggaaattc	28800
cggagttaga	aggtagaata	actaaaattt	aaaatttact	agagaagggt	caacactata	28860
tttgaactgg	cagaagaaaa	athtagtgag	acaaatatac	ttcaatagac	attattcaaa	28920
tgaaaaataa	aaagaaaaaa	gaatgaagaa	aaataaacag	aatctcagca	aaatgtggca	28980
ccacattaat	cacattaaca	tatgcatact	gagagtaccg	gaagcagatg	agaaaaggga	29040
agaaaaata	ttcaaatgat	ggccagtaac	ttcctagatt	tttgttttaa	agcaataacc	29100
tatacaatca	agaaactcaa	tgaattccaa	gtaggataaa	tacaaaaaga	accacaaa	29160
gatacaaccat	ggtaaaaatg	ctgtaagtca	aaaacagaga	aaatattgaa	agcagctaga	29220
ggaaaactta	taagagaacc	tcacttacaa	aagaacatca	cttataaaaag	aaccacaata	29280
atagaaaacag	ttgacctctc	atcagaaaaca	atgaatgata	acataatttga	agtgtcctaaa	29340
gaaaaaaaat	aaagattcct	atatacgaca	aagctgtctt	tcaaaaaatat	acatccaaaa	29400
ggattgaaac	caggggtcctg	aagagttatt	tgtacatcca	tgttcatagc	agcattatct	29460
acaatagcca	aaaggtagaa	gcaacccaag	ggtccatcga	caaataaata	aaatgtggta	29520
tatgtataca	caatgggaatt	tattcagtat	taaaaaggaa	tgaattctcg	acacatgcta	29580
caacatggct	aaaccttgag	aacactatgc	taagtgaat	aagccagcca	caaaaggaca	29640
aataccatata	tacttcaact	gtatgaaaata	cctagggtag	tcaaattcag	agatagaaaag	29700
taaaacagtg	gttgccaagg	gctgagggag	ggagtaacgt	ggagttattg	tgaatgggt	29760
acagaatttc	agttttgcaa	gataaaaaaa	gttctggaga	cagatgggtgg	tgaggggtgg	29820
acaacaatac	aaatatactt	tatactactg	aacagtatatac	ttaaaaatga	ttaacattggt	29880
gaaaccccg	ctctactaaa	aatacaaaaa	aattagctgg	gtgtgggtggc	gggcacctgt	29940
aatccagct	acttgggagg	ctgagggcagc	agaattgctt	gaaaccagaa	ggcggaggtt	30000
gcagtgaact	gagattgcgc	caccgcactc	tagcctgggc	aataagagca	aaactccgtc	30060
tcaaaaaata	aaaaataaaa	aaaattttaa	aatgattaa	caggaggcca	ggcaggggtg	30120
ctcacaccta	taatgccagc	actttgggag	gccgaggcag	gcgatcactt	gagaccagga	30180
gtttgagacc	agcctggcca	acatggcaaa	accctgtctc	tgctaaaaat	acaaaaatta	30240
gccaggcatg	gtggcatata	cttataatcc	cagctactgg	tgagactgag	acacgagaat	30300
tgcttgaacc	caggaggcag	agattgcagt	gagtcgagat	cgcgccactg	aattccagcc	30360
tggggcgacag	agcaagattc	tgtctcgaaa	aaacaaaaac	aaaaacaaaa	agcaaaaacca	30420
aaaaataa	aagcaggaaa	cgagattgct	gctgaggagg	agaaagatgt	gcaggaccaa	30480
ggctcatgag	agcacaacaa	ttttcaaaaa	atgtttaatg	attaaaaatg	taaaatttat	30540
atgtatctta	ccacaaaaaa	aagggtctgg	gggcaggaaa	tgaagggtgaa	ataaagacat	30600
cccagagaaa	caaaagtaga	gaatttgttg	ccttagaaga	aacaccacag	gaagttcttc	30660
agggtgaaaa	caagtgaccc	cagagggtaa	tctgaattct	cacagaaaaat	tgaagcatag	30720
cagtaaaagg	tattctgtaa	ctatgacact	aacaatgcat	atttttctct	ttcttctctg	30780
aaatgatatt	aaaagcaatt	gcataaaaata	ttatatataa	agcctattgt	tgaacctata	30840
acatatatag	aaatatactt	gtaatatatt	tgcaaaataac	tgcaaaaaag	agagttggaa	30900
caaagctgtt	actaggctaa	agaaattact	acagatagta	aagtaataata	acaggggaact	30960
taaaaaataa	attttaaaaa	atttaaaaaat	aataattaca	acaataatat	ggttgggttt	31020
gtaatatata	tagacataat	acaaaaatac	cacaaaaagg	gaagaagaca	atagaactac	31080

25/122

ataggaataa	catttttggt	tctaactaga	attaaattat	aaatatgaag	tatatctctgg	31140
taagttaaga	cacacatgtt	aaaccctaga	tactaaaaag	taactcacat	aaatacagta	31200
aaaaataaaa	taaaataatt	aaaatgtttg	tattagtttc	ctcaggggtac	agtaacaaac	31260
taccacaaat	tgagtggctt	aacacaactt	aaatgtat	tctcccagtt	ctggaggcta	31320
aacacctgca	atcaagggtga	gtacagggcc	atgctccctg	tgaaggctct	aggaaagat	31380
cctcccttgt	ctctccagc	ttccagtggt	tctcagtaac	cctaagtgt	ccttggttg	31440
tagctatata	attcctagca	accagaaaaga	agaaaataat	aaagattatg	gcaaaaaata	31500
atgaaatcaa	aaggagaaaa	atggaaaaaa	ataaataaaa	ccaaaagcta	gttctttgaa	31560
aagatcaacc	aagttaacaa	accttttaac	tagactgaca	aaaaggaggt	aagactcaaa	31620
ttactagaat	cagaaataaa	agagggggaca	ttactaatga	gggattagaa	aagaatacta	31680
cgaacaaatg	tgtgccaaca	aattagaaaa	cttagatgaa	atggacaggt	tcctaggaca	31740
acatcaacta	ccaaaattta	ctcaagaaga	aagagacaat	ttgaatgagc	tataacaagg	31800
gaagagactg	aattgacaac	caagaaacta	tccacaaaga	aaatcccagg	cccagaagat	31860
ttcactgtga	aattctttca	aacttataaa	tataaattaa	catcagttct	tcacaaactc	31920
ctccaaaaaa	aagaacagat	ctctatttac	aggcgatacg	atcttttagaa	aatcctaagg	31980
gaactactaa	gacactatga	taactgataa	acaagttcag	caaggctgca	ggatagaaaa	32040
ccatataaca	aaaatctatt	atattttctat	acactttgcag	tgaacaaccc	aaaaatgaga	32100
tttagaaaat	aattcaattt	acaataacat	caaaaagaat	aaaaacactc	aaaaataaat	32160
ttattcaagt	aagtgcataa	cttataactct	agaagctaca	aaacactgtt	aaaagaaatt	32220
aaaggtttac	ataaatgaaa	aactatccca	tgttcatgga	tcaaaagact	tattactggc	32280
atgctctccc	aaattgatct	ataaattcaa	caaaatcctt	atcaaaatcc	cagatgaggg	32340
tgggggtggc	ggttcatgcc	tgtaatccca	gcactttggg	aggctgaggg	acgcagatta	32400
cctgaggtcg	ggagctcgag	atcagcctga	ccaacatgga	gaaaccctat	ctcttctaaa	32460
aatcaaaaat	tagtcaggcg	tgggtggcaca	tgcctataat	cccagctact	cggggaagctg	32520
aggcaggaga	atcgcttgaa	cccaggaggg	agaggttgca	gtgagccaag	atcgtgccat	32580
tgcactccag	cctgggcaac	aagagcaaaa	ttccatctca	aaaaaaaaaa	aaaaaaaaatc	32640
ccagatgact	tcactgttga	aattgaaaaag	attattctaa	aattcacatg	gaattgcaag	32700
accttgagaa	tagccaaaac	aaacttgaaa	aacacgaaca	aaatatagga	tgactcactt	32760
gccaattgca	aatgttacga	cacagcaaca	gtaattcaaga	ctgtgtggta	ctggcaaaag	32820
acacatacat	acatacatat	caatggaata	taattgagag	tacagaaaaca	agcctaataa	32880
tctatggtaa	gtgcttttct	atTTTTTct	TTTTTTTTT	ctTTTTTgta	gagatagaat	32940
ctcaccatgt	tgcccaggct	ggtcttcaac	ttctgggctc	aagcaatcct	cccactgtgg	33000
cctcccaaag	tgctgggata	actggcatga	gccaccacat	ccagcccaga	tgattttcaa	33060
aaaagtcaac	aagaccattc	ttttcaacaa	ataggtctgg	gatgatcaga	tagtcacatg	33120
aaaaaaaaaa	tgaagttgga	ccctccatca	cactaaagtg	ctgcgattat	aggcatcagc	33180
caccacatcc	agcccaaattg	atTTTcaaaa	aggtcaacaa	gaccattctt	ttcaacaaat	33240
aggctctggga	taatcagata	gtcacatgaa	aaaaaaaaatg	aagttggacc	ctccatcaca	33300
ccatattgcaa	aaattaattc	aaaaatgaat	tgatgactta	aacgtaaagag	ttacgactgt	33360
aaaactctta	gaaggaaaaca	tacgggtaaa	tcttaaagac	gttaggtttg	acaaagaatt	33420
cttagacatg	acaccaaag	catgaccaac	taaggtaaaa	taggggtaaa	tgtacctacc	33480
aaaatgaaaa	acctttgtgc	tggaaaaggac	accatcaaga	aatggaaaagc	caaaatagcc	33540
aaggcaatat	taagcaaaaa	gaacaaagct	ggaggcatca	tactacctga	cttcaaagca	33600
acagtaacca	aaacagcatg	gtactagtag	aaaaacagac	acatagacca	atggaacaga	33660
ataaagaacc	caaaaataaa	tccacatatt	tatagtcaac	tgatttttga	caatgacacc	33720
ccttcaataa	atgatactag	gaaaactgga	tatcgatatg	cagaagaata	aaactagacc	33780
cctatctctc	accatataga	aaaatcaact	cagactggaat	taaagacttg	aatgtaagac	33840
ccaaaactat	aaaactactg	gtagaaaaca	taaggaaaaa	cgcttcagga	cattgggtcca	33900
ggcaaagatc	ttatggctaa	aacctcaaaa	acacaggcaa	caaaaacaaa	aatggaaaaa	33960
tagcacttta	ttaaactaaa	aagctcctgc	acagcaaaag	aaacaacaga	atgaaaagac	34020
aacctgtaga	atgggagaaa	atatttgcaa	actatccatc	catcaaggga	ctagtatcca	34080
gaacacacaa	gtgactaaaa	caactcaaca	gcaaaaaagc	aaataatctg	gtttttatat	34140
gggcaaaaaga	tctgaataaa	cattctcaaa	ggaagacata	caaatgtcac	tatcattctg	34200
ccagtaccac	actgtcttga	ttacttgtta	gtgtataaat	ttttaaattg	ggaagtgtga	34260
gtcatcctac	actttgttct	tgTTTTTcaa	gtttgttttg	gctattctgg	gagccttgca	34320
agtataaaaat	agccaacaag	tatgaaaaaa	gtctcaccat	cactaatcat	cagagaaata	34380
aaaaatcaaga	ccactatgag	atatcctctc	actccagtta	gaatggctac	tatcaaaaaag	34440
acaaaatata	atggatgctg	gcaaagattt	ggagaaaagg	gaactcctat	acactgtggg	34500
tagggatgca	aattggtaat	ggccattatg	gaaaataata	ctgaggtttt	tcaaaaaact	34560
gaaaatagaa	ctaccatatg	atccagcaac	cctactactg	ggtatttatc	caaaggaaag	34620
aagtcagtat	actgaagaaa	tatatgcact	ctcatgttaa	ttgcaacact	gttcacaaca	34680

26/122

gccaagacag	ggaataaatc	taaatgtgca	tcaacagatg	aatggataaa	gaaaatgtgg	34740
catatacact	caatagaata	ctattcagcc	attaaagaag	aatgaaatcc	tgatcatccca	34800
gcaacatgga	tgaacctgga	ggacattata	tttaattgaaa	taagtaaagc	acaaaaagat	34860
aaacagtaca	tggtctcact	cagacatggg	tgctaaaaag	aaaatggggg	cacagaatta	34920
gaagggggagg	cttggggaaaa	gttaatggat	aaaaattttac	agctatgtaa	gaagaataag	34980
tttttagtggt	ctatagaact	gtagggcgag	tatagttacc	aataacttat	tgtacatggt	35040
caaaaaagcta	gaagagattt	tggatgttcc	cagcaciaag	gaatgataaa	tggtttgtgat	35100
gatggatatac	ctaattaccc	tgattcaatc	attacacatt	gcatacatgt	atcaaatat	35160
cactctgtac	ctcataaata	tgtataatta	ttacgtcaac	aaaaaaaagg	aaaaaaaagaa	35220
aattaagaca	acccacataa	tggaaagaaat	aaaatatctg	caaattatat	atatctgata	35280
aatatttaaat	atttataata	tataaagaac	tcctacaact	caagaacaac	aacaaaacaa	35340
cccaattcaa	aaatgggtaa	aagccttgaa	tatacactta	tctaaagact	atatacaatt	35400
ggccaataaaa	gacacgaaaa	gatgctcaac	atcactagtc	atcagggaaa	tataaatcaa	35460
aaccacaatg	tagaatgtag	acaccacttc	atatgcacta	ggatggctag	aataaaaagg	35520
taataacaaa	tggtggtaag	gatgtgaaaa	aatcagaaac	ctcattcgct	gctgttgagg	35580
atgtaaagt	atgcagccac	tttggaaaa	agtctggcag	ctcctcaaat	tattaaatac	35640
agagttaccg	tatgacccag	gaatattcct	cctgggtcta	taacaaaaaa	aatgaaaaca	35700
tatatccaca	taaaaacttg	tacatgggca	tttatagcaa	cattattcat	aacagcaaa	35760
gtggtaagaa	cccatatgcc	catcatctga	tgaacaggta	aataacatgc	ggtattatcc	35820
atacactaga	atattatctg	cccatacaag	gagtgacatc	cagctacatg	ctacaaggat	35880
gaatctcgga	aaccttatgc	taagtgaag	aagccagtca	caaatgacca	cagattatga	35940
ttccatgcat	cggaaatgac	cagaataggg	aaatctatag	agacagaaag	tagattagt	36000
gttgggtggg	gctgggagga	caggtagtac	actactttcc	cagaactact	ggaacaaagt	36060
accacaaact	ggggagctta	aacatagaaa	ttgatttctc	cacagttctg	gagactagga	36120
ctctgagatc	aaggtgtcag	cagagctggg	tctttctgag	ggccctgagg	caaggctctg	36180
tcccaggcct	ctctccttgg	ctggcagggt	gccatcttct	ccctgcgtct	tcacatcatc	36240
ttttctctgt	gtgtgcccat	gtccaaattt	tgattggctc	attctgggtc	atggccaatt	36300
gctatgcaca	aagtgaagtc	tacttccaaa	agaagggaa	agggaaacact	gactaggcta	36360
aacttatagt	cattttaatg	tccgcttttc	ctatgagatt	gtgaacacac	agaagtaggg	36420
tttttatcta	cattgtgcaa	agtttaataa	gaaaaataga	attcaagaga	agcagttcaa	36480
tagcaggaat	ttaatatggg	aactaattac	aaggtttagg	gcaggactaa	aaagccagtt	36540
gggatgggta	gccaacccag	agattagcaa	cagtgaggac	ccatctacct	accaccatg	36600
aagctggaag	gataaaggag	gggctattat	cagagtcac	aagccagtgt	cagagtcctt	36660
ggctggagct	gggaccaccc	tagagacact	gtgcaaagca	gaaaacaagg	gggaaaaacc	36720
ctgacttctc	ccttctctcc	accttcaat	ctcccactag	tgcttctctac	tagccatact	36780
tggccagaga	cagtgacaag	gaacactgca	aaatgaagtt	tgtaggaatc	atctccctct	36840
gagacagaga	aatatggaag	ggtagaaaat	gaatcagagg	ataaagagaa	aaaaccctga	36900
gtactatctt	atttatcttt	gtatctccag	tgcctaattct	gtctctcaaa	aaaggaaagc	36960
aattgagaga	aactgaaaac	tccaattgaa	atgaaagaat	ggagaattac	tggactagaa	37020
gagaagagaa	aaattttattc	cgcataaggt	aaacaagaat	ggattcacaa	aggacgtgat	37080
gaatgaaaag	ctataatcag	caaagatttg	ccagagaaat	taaaaagtgg	taaactcagc	37140
cacgctgtac	aacctgaagg	cacaatgcat	gaaaacgttt	caagaaatga	caagatttga	37200
agtcaaattc	taagtgtctt	tccagaatct	ctcaagacga	ttatatagct	accccatctt	37260
attaaataaaa	atggaaaactt	actaaaactt	ccccttgat	taaaactaaca	tatgtcctaa	37320
tagcaaacga	ttctgggaatt	cctagagtaa	aatatatttc	gtcaaagtgt	attgctcttt	37380
taatattctg	ctgacctcct	tttgctattt	aggatatttg	tatacacatc	acacgtaaat	37440
ttgggtctata	gtttacatct	acgggcttat	actgttcttt	ttttcatctt	tttaaaattt	37500
ccaaccccca	gtatccatat	actgctctct	atcaggggta	ttttaacttt	gtaaaatcag	37560
ctgagatgct	ttccatgttt	ttttttttta	ttttctgcca	catttgaata	gcataggagt	37620
taccaccatc	aaccttggat	tatttaagca	ttcacgattc	cacgtgtgga	ttttttattc	37680
agagtctttc	ttgtcattcc	tgctatcagc	acagaaccca	atctcagctt	tccagctata	37740
ctctcaccoc	atgggaatttg	cagatgaagt	tcaaaaggac	ctttgcatta	tcctgcctcg	37800
ccctctctccc	ccttcattta	gacatcacct	tcttctagaa	cgtcttacct	gacatgccct	37860
gctcccaacc	actgtgtgccc	aattgtgtgc	tctcccggtg	cctggcctgc	catcctcttt	37920
agtaattgcc	tgctccctca	tctgtctccc	caccagaca	ttaagctgaa	tagactggat	37980
ttgtgtcttg	tccatcacta	taatctcagc	acctagtacc	tagtaggtac	ttacctgta	38040
ttcattagca	aaatgttatg	tataaccttg	caccttaaaa	acaagagaag	gaagacaaaa	38100
ttaagtctta	agactatggt	ttagaacatg	gatcagaaac	tacagtctgc	agcccaaatc	38160
cagaccaaat	gaagagacca	tggtcattta	catacaacct	atagcagctt	tcacactaca	38220
ggagcagagc	taagtagttc	caagggaaca	cacggccctg	caaagcctaa	aatattttact	38280